The use of heuristic optimization algorithms to facilitate maximum simulated likelihood estimation of random parameter logit models

Arne Risa Hole[†] and Hong Il Yoo[‡]

[†]Department of Economics, University of Sheffield
a.r.hole@sheffield.ac.uk

[‡]Durham University Business School, Durham University
h.i.yoo@durham.ac.uk

November, 2016

Abstract

Applications of random parameter logit models can be found in various disciplines. These models have non-concave simulated likelihood functions and the choice of starting values is therefore crucial to avoid convergence at an inferior optimum. Little guidance exists, however, on how to obtain good starting values. We apply an estimation strategy which makes joint use of heuristic global search routines and gradient-based algorithms. The central idea is to use heuristic routines to locate a starting point which is likely to be close to the global maximum, and then to use gradient-based algorithms to refine this point further. Using four different empirical data sets, as well as simulated data, we find that the proposed strategy locates higher maxima than more conventional estimation strategies.

Keywords: mixed logit, generalized multinomial logit, differential evolution, particle swarm optimization

1 Introduction

With an increase in desktop computing power, the random parameter logit model (RPL) has become increasingly common in empirical applications. Also known as mixed logit, RPL provides a flexible framework for modeling discrete choice data. RPL can approximate any random utility maximization model arbitrarily well subject to specifying a suitable joint distribution of parameters (McFadden and Train, 2000), and incorporate preference heterogeneity between different individuals alongside panel correlation across observations on the same individual (Revelt and Train, 1998). Applications of RPL can be found in a range of disciplines including economics, marketing science, transportation studies and health services research.

While RPL is specified by augmenting the parameters of the multinomial logit model (MNL) with random heterogeneity, RPL poses a number of estimation issues which MNL does not. Perhaps the best known one is that in most applications, the RPL likelihood is a multidimensional integral which has no closed-form expression and needs to be numerically approximated by using simulation. This issue has motivated several studies to explore how best to obtain a more accurate approximation from a given number of draws from the joint distribution of random parameters (Train, 2009, pp.205-236), and their findings have popularized the use of Halton sequences to generate draws. While progress has also been made on developing estimation methods which are more computationally attractive than the classical method of maximum simulated likelihood (MSL) in certain aspects (Huber and Train, 2001; Harding and Hausman, 2007; Train, 2008), MSL still remains the most commonly used method as it can be readily applied in conjunction with almost any joint distribution of random parameters.

This paper applies an estimation strategy to address another well-known estimation issue, on which limited practical guidance exists. Specifically, in contrast to its MNL counterpart, the RPL likelihood is not globally concave and may feature several local maxima. As in other similar contexts of non-linear estimation, the selection of "good" starting values for estimated parameters is crucial to avoiding an inferior local maximum. In the RPL literature, nevertheless, empirical studies rarely provide an explicit discussion of starting values used, and the question of how to obtain "good" starting values has not been the subject of inquiry as far as we know.

Our proposed estimation strategy makes joint use of heuristic optimization algorithms and usual gradient-based algorithms to obtain the MSL estimates of RPL. Following Dorsey and Mayer (1995), the central idea is to use the heuristic algorithms to locate a starting point which is likely to be close to the global maximum, and then to use gradient-based algorithms to refine this point further. For the heuristic search step, we consider two parsimonious but effective algorithms which can be easily implemented by non-specialists in heuristic optimization: the differential evolution (DE) algorithm (Storn and Price, 1997) and the particle swarm optimization (PSO) algorithm (Eberhart and Kennedy, 1995). These population-based algorithms are well-suited to the task of locating candidate solutions away from inferior maxima, as they search comprehensively over the parametric space in looking for the directions of improvement (Gilli and Winker, 2009). As other gradient-free algorithms, however, they tend to be much slower than gradient-based algorithms in refining a candidate solution to a nearby maximum. Our estimation strategy exploits the global search efficiency of the population-based heuristics and the local search efficiency of gradient-based algorithms, in the sense of Dorsey and Mayer (1995).

We provide computational evidence on the performance of the DE- and PSO-assisted estimation strategies in four different empirical data sets of varied sizes, as well as in simulated data sets. While these strategies can be applied to the estimation of any RPL specification, the case studies primarily focus on the generalized multinomial logit model (GMNL) of Fiebig et al. (2010). The results suggest that the DE-assisted strategy is a very effective tool to diagnose whether a solution obtained by following the conventional practice is a global maximum. In all four empirical data sets, the DE-assisted strategy locates solutions which improve on the best conventionally obtained solutions in terms of maximized log-likelihood. Under most computational settings improved solutions are found with high enough empirical frequencies to suggest that a small number of DE-assisted estimation runs would be sufficient for detecting whether a preferred conventional solution is at an inferior maximum. While the PSO-assisted strategy also locates solutions improving on the best conventional solutions in all four empirical data sets, it does so with much lower empirical frequencies. Moreover, in each data set, the best solution that attains the highest likelihood we have found comes from the DE-assisted strategy. The findings from simulated data sets support these results.

The remainder of this paper is organized as follows. Section 2 reviews the specification and MSL estimation of GMNL. Section 3 presents the DE and PSO algorithms. Section 4 presents the main case studies based on two smaller empirical data sets. Section 5 briefly introduces the further case studies reported in the Online Appendix, which explore the applicability of the findings to two larger empirical data sets, simulated data sets and other computational settings. Section 6 concludes.

2 The generalized multinomial logit model

We assume a sample of N individuals who make a choice from J alternatives in each of T choice situations. The utility person n derives from choosing alternative j in choice situation t is specified as

$$U_{njt} = \boldsymbol{x}'_{njt}\boldsymbol{\beta}_n + \varepsilon_{njt} \tag{1}$$

where $\boldsymbol{x}_{njt}$ is an L-vector of alternative attributes, $\boldsymbol{\beta}_n$ is a conformable vector of utility coefficients, and ε_{njt} is an idiosyncratic error term which is independent and identically distributed as type 1 extreme value. Specifying a non-degenerate density of $\boldsymbol{\beta}_n$ leads to a random parameter logit model (RPL), which allows for interpersonal heterogeneity in preferences for variations in different attributes (Revelt and Train, 1998; McFadden and Train, 2000).

In the generalized multinomial logit model (GMNL) of Fiebig et al. (2010), $\boldsymbol{\beta}_n$ is specified as

$$\boldsymbol{\beta}_n = \mu_n\boldsymbol{\beta} + \{\gamma + \mu_n(1-\gamma)\}\boldsymbol{\eta}_n \tag{2}$$

where scalar γ and vector $\boldsymbol{\beta}$ are deterministic, and random vector $\boldsymbol{\eta}_n$ is distributed $MVN(\mathbf{0}, \boldsymbol{\Sigma})$. Using $\boldsymbol{z}_n$ to denote an M-vector of individual n's characteristics, the random scale factor μ_n is further specified as

$$\mu_n = \exp(\overline{\mu} + \boldsymbol{z}'_n\boldsymbol{\theta} + \tau v_n) \tag{3}$$

where scalar τ and vector $\boldsymbol{\theta}$ are deterministic, and random scalar v_n is distributed $N(0,1)$. Scalar $\overline{\mu}$ is a normalizing constant which is calibrated to set the mean of μ_n to 1 when $\boldsymbol{\theta} = \mathbf{0}$. This model can be interpreted as one that accommodates both canonical "coefficient heterogeneity" through individual-specific deviations $\boldsymbol{\eta}_n$ around population mean coefficients $\boldsymbol{\beta}$, and "scale heterogeneity" through the individual-specific scale factor μ_n. Its flexibility is enhanced by the γ parameter which lets scale heterogeneity affect $\boldsymbol{\beta}$ and $\boldsymbol{\eta}_n$ differently.

Conceptually, allowing the scale factor μ_n to vary by n can be motivated by the possibility that some individuals make choices which are "noisier", or less aligned with variations in the observed attributes, than others. Then, the idiosyncratic unobservables ε_{njt} would have a larger variance for those individuals, making the scale factor smaller.[1] As can be seen from equation (2), however, scale heterogeneity is equivalent

[1]This directly follows from the usual identification result for discrete choice models that when ε_{njt}

to a particular type of coefficient heterogeneity, so the two cannot be sharply distinguished from each other (Fiebig et al., 2010, p.398). The main empirical attraction of GMNL is that the random parameter specification in (2) can approximate a wide range of preference patterns, some of which would otherwise call for the use of much less tractable specifications (Keane and Wasi, 2013).

Several other discrete choice models can be derived as special cases of GMNL. The GMNL-I and GMNL-II models (Fiebig et al., 2010) are obtained by setting γ to 1 and 0, respectively. The GMNL model reduces to the standard mixed logit model when the scale factor is assumed to be constant ($\mu_n = 1$), while the the MNL model with scale heterogeneity (SMNL) is obtained by constraining the covariance matrix of $\boldsymbol{\eta}_n$, $\boldsymbol{\Sigma}$, to $\mathbf{0}$. If both of these constraints are imposed simultaneously, the standard multinomial logit model is obtained. The various special cases of GMNL are summarized below:

- GMNL-I: $\boldsymbol{\beta}_n = \mu_n \boldsymbol{\beta} + \boldsymbol{\eta}_n$ ($\gamma = 1$)
- GMNL-II: $\boldsymbol{\beta}_n = \mu_n(\boldsymbol{\beta} + \boldsymbol{\eta}_n)$ ($\gamma = 0$)
- SMNL: $\boldsymbol{\beta}_n = \mu_n \boldsymbol{\beta}$ ($\boldsymbol{\Sigma} = \mathbf{0}$)
- Standard mixed logit (MIXL): $\boldsymbol{\beta}_n = \boldsymbol{\beta} + \boldsymbol{\eta}_n$ ($\mu_n = 1$)
- Standard multinomial logit (MNL): $\boldsymbol{\beta}_n = \boldsymbol{\beta}$ ($\mu_n = 1$ and $\boldsymbol{\Sigma} = \mathbf{0}$)

The probability that individual n makes a particular sequence of choices is given by:

$$S_n = \int \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\boldsymbol{x}'_{njt} \boldsymbol{\beta}_n)}{\sum_{j=1}^{J} \exp(\boldsymbol{x}'_{njt} \boldsymbol{\beta}_n)} \right]^{y_{njt}} f(\boldsymbol{\beta}_n | \boldsymbol{\beta}, \gamma, \tau, \boldsymbol{\theta}, \boldsymbol{\Sigma}) d\boldsymbol{\beta}_n \tag{4}$$

where $y_{njt} = 1$ if the individual chose alternative j in choice situation t and 0 otherwise and density $f(\boldsymbol{\beta}_n | \boldsymbol{\beta}, \gamma, \tau, \boldsymbol{\theta}, \boldsymbol{\Sigma})$ is implied by equation (2). The parameters $\boldsymbol{\omega} = (\boldsymbol{\beta}, \gamma, \tau, \boldsymbol{\theta}, \boldsymbol{\Sigma})$ can be estimated by maximizing the simulated log-likelihood function

$$SLL(\boldsymbol{\omega}) = \sum_{n=1}^{N} \ln \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(\boldsymbol{x}'_{njt} \boldsymbol{\beta}_n^{[r]})}{\sum_{j=1}^{J} \exp(\boldsymbol{x}'_{njt} \boldsymbol{\beta}_n^{[r]})} \right]^{y_{njt}} \right\} \tag{5}$$

where $\boldsymbol{\beta}_n^{[r]}$ is the r-th draw from the density of $\boldsymbol{\beta}_n$ and R is the total number of draws.

is normalized as an iid variable, the overall scale of utility is inversely related to the true idiosyncratic variance.

The standard approach to maximizing the simulated log-likelihood function is to use a gradient-based method such as the Newton-Raphson or Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithms. See Train (2009, pp.185-204) among others for a description of these methods. The researcher starts with an initial guess of the solution - the starting values - which are then improved upon by the algorithm until a specified stopping criterion is reached. As is well-known, gradient-based methods cannot distinguish between local and global maxima, and will declare convergence if either type of maximum is reached. Thus, unless the function to be optimized is globally concave, it is not guaranteed that the solution is the global maximum. This issue is of practical importance since the simulated log-likelihood function of the GMNL model and its special cases (with the exception of the MNL model) is not globally concave, much as that of other RPL models. In particular, different starting values may lead to different solutions, which suggests that applied researchers should try different sets of starting values to investigate how sensitive the results are to the particular values used. The choice of starting values is rarely discussed in applications of GMNL and other RPL models, however. We present some of the strategies that researchers may employ in the following section.

3 Population-based optimization heuristics

This section describes alternative estimation strategies which use population-based heuristic optimization algorithms to obtain starting values for the gradient-based methods. We focus on two population-based optimization heuristics, namely the differential evolution (DE) algorithm (Storn and Price, 1997) and the particle swarm optimization (PSO) algorithm (Eberhart and Kennedy, 1995), which have been found to outperform many other heuristic algorithms in a wide range of applications (Gilli and Winker, 2009; Das and Suganthan, 2011).

The main operational aspects of these algorithms are as follows. Suppose that there are a total of K parameters in $(\boldsymbol{\beta}, \gamma, \tau, \boldsymbol{\theta}, \boldsymbol{\Sigma})$ and let a candidate solution be the K-vector of guesses about those parameters. Each algorithm is initialized by generating $\texttt{P}$ different random starting points forming the initial "population" of candidate solutions, where $\texttt{P}$ is a large number. Then, every one of these candidate solutions is updated over $\texttt{G}$ iterations, or "generations", where $\texttt{G}$ is another large number. Within each generation, the rule for updating each solution takes into consideration the population of solutions at the end of the preceding generation. The rule also features random elements influencing the direction and extent to which each solution gets updated. In the end, the terminal population of P candidate solutions are obtained, and the best candidate solution in the sense of giving the highest simulated log-likelihood value is selected as the fully iterated solution.

For further discussion, let $\boldsymbol{\omega}^{g,p} = (\boldsymbol{\beta}^{g,p}, \gamma^{g,p}, \tau^{g,p}, \boldsymbol{\theta}^{g,p}, \boldsymbol{\Sigma}^{g,p})$ denote a K-vector of possible values of model parameters. Superscripts $p = 1, 2, \cdots, \mathtt{P}-1, \mathtt{P}$ and g =0,1,$\cdots$, $\mathtt{G}-1, \mathtt{G}$ identify the pth candidate solution at generation g. Let $\boldsymbol{\Omega}^g = (\boldsymbol{\omega}^{g,1}, \boldsymbol{\omega}^{g,2}, \cdots, \boldsymbol{\omega}^{g,\mathtt{P}-1}, \boldsymbol{\omega}^{g,\mathtt{P}})$ be the collection of P up-to-date candidate solutions as at g. For later use, we define $g' \equiv g - 1$.

Once the initial population $\boldsymbol{\Omega}^0$ has been generated, each algorithm can be implemented by setting up a simple loop as follows:

```
for g = 1 to G {
    for p = 1 to P {
        DE^{g,p}(F, Cr) or PSO^{g,p}(C, D)
    }
}
```

DE^{g,p}(F, Cr) and PSO^{g,p}(C, D) are the rules that the respective algorithms apply to compute the updated candidate solution $\boldsymbol{\omega}^{g,p}$. Each rule depends on two "tuning parameters" (F, Cr) or (C, D), which are user-specified scalar inputs much as the population size P and the number of generations G. We now turn to a more specific description of each rule.

3.1 Updating process under differential evolution (DE)

The updating rule DE^{g,p}(F, Cr) consists of three main stages: mutation, recombination and selection. The first two stages produce a K-vector of trial values $\boldsymbol{t}^{g,p}$. This is competed against $\boldsymbol{\omega}^{g',p}$ in the last stage, which selects the better of the two vectors as $\boldsymbol{\omega}^{g,p}$.

The mutation stage uses the amplification factor F and constructs a linear combination of three existing candidate solutions other than $\boldsymbol{\omega}^{g',p}$. To this end, three vectors are randomly drawn from $\boldsymbol{\Omega}^{g'}\backslash\{\boldsymbol{\omega}^{g',p}\}$ with equal probabilities and without replacement: let these draws be $\boldsymbol{\omega}^{g',z_1}, \boldsymbol{\omega}^{g',z_2}$ and $\boldsymbol{\omega}^{g',z_3}$. Their linear combination $\boldsymbol{d}^{g,p}$ is specified as

$$\boldsymbol{d}^{g,p} = \boldsymbol{\omega}^{g',z_1} + \mathtt{F}(\boldsymbol{\omega}^{g',z_2} - \boldsymbol{\omega}^{g',z_3}). \tag{6}$$

The recombination stage uses the cross-over probability $\texttt{Cr}$ to construct the K-vector $\boldsymbol{t}^{g,p}$ by combining elements of $\boldsymbol{\omega}^{g',p}$ and $\boldsymbol{d}^{g,p}$. This step also involves making $K+1$ different random draws: a positive integer $i^{g,p}$ is drawn from $\{1, 2, \cdots, K-1, K\}$, while K scalars $u_k^{g,p}$ for $k = 1, 2, \cdots, K-1, K$ are drawn from the standard uniform distribution. Now, let $\omega_k^{g',p}$,$d_k^{g,p}$ and $t_k^{g,p}$ denote the kth elements of $\boldsymbol{\omega}^{g',p}$,$\boldsymbol{d}^{g,p}$, and $\boldsymbol{t}^{g,p}$ respectively. Each element of $\boldsymbol{t}^{g,p}$ is chosen according to the following criteria:

$$\begin{aligned} t_k^{g,p} &= d_k^{g,p} \text{ if } u_k^{g,p} \leq \texttt{Cr} \text{ or } k = i^{g,p} \\ t_k^{g,p} &= \omega_k^{g',p} \text{ otherwise} \end{aligned} \tag{7}$$

Due to the role of integer $i^{g,p}$, $\boldsymbol{t}^{g,p}$ is always different from $\boldsymbol{\omega}^{g',p}$ in at least one element.

The selection stage evaluates the simulated log-likelihood (5) at the updating target $\boldsymbol{\omega}^{g',p}$ and at the trial vector $\boldsymbol{t}^{g,p}$. The updated solution $\boldsymbol{\omega}^{g,p}$ equals $\boldsymbol{t}^{g,p}$ if $SLL(\boldsymbol{t}^{g,p}) > SLL(\boldsymbol{\omega}^{g',p})$, and $\boldsymbol{\omega}^{g',p}$ otherwise. The terminal population $\boldsymbol{\Omega}^G$ consists of $\texttt{P}$ candidate solutions which have thus been updated $\texttt{G}$ times. It is the best solution in $\boldsymbol{\Omega}^G$, in the sense of giving the highest simulated log-likelihood, that is passed to a gradient-based algorithm for further improvement.

The role of the amplification factor $\texttt{F}$ can be likened to that of the step size in gradient-based optimization. In the above updating rule, $\texttt{F}$ is the only component that can be systematically increased by the user to induce a large extent of parametric changes between generations. The cross-over probability $\texttt{Cr}$, on the other hand, influences how often the parametric changes are finalized. Storn and Price (1997) find in a range of applications that while $\texttt{F}$ is not a probability, the DE algorithm tends to perform the best when it is chosen from the $(0, 1)$ interval much as $\texttt{Cr}$.

3.2 Updating process under particle-swarm optimization (PSO)

The updating rule $\text{PSO}^{g,p}(\texttt{C},\texttt{D})$ deviates from $\text{DE}^{g,p}(\texttt{F},\texttt{Cr})$ in that now $\boldsymbol{\omega}^{g,p}$ always changes from $\boldsymbol{\omega}^{g',p}$ even when doing so worsens the simulated log-likelihood. Two additional concepts are needed for a further exposition. First, define $\boldsymbol{s}^{g,p}$ as the best pth candidate solution that has been obtained up to generation g: that is, $\boldsymbol{s}^{g,p}$ is the best one out of $\boldsymbol{\omega}^{0,p}, \boldsymbol{\omega}^{1,p}, \cdots, \boldsymbol{\omega}^{g-1,p}, \boldsymbol{\omega}^{g,p}$. Likewise, define $\boldsymbol{q}^g$ as the best candidate solution that has been obtained up to generation g: that is, the best one of out $\boldsymbol{s}^{g,1}, \boldsymbol{s}^{g,2}, \cdots, \boldsymbol{s}^{g,p-1}, \boldsymbol{s}^{g,p}$.

$\text{PSO}^{g,p}(\texttt{C},\texttt{D})$ uses the acceleration constant $\texttt{C}$ and the inertia weight $\texttt{D}$ to "fly" $\boldsymbol{\omega}^{g',p}$

towards the best-so-far positions at $\boldsymbol{s}^{g',p}$ and $\boldsymbol{q}^{g'}$, thereby obtaining the updated solution $\boldsymbol{\omega}^{g,p}$. The extent of the involved changes, or "velocity of the flight" $\boldsymbol{v}^{g,p}$, depends also on two scalars $r_1^{g,p}$ and $r_2^{g,p}$, each of which is drawn from the standard uniform distribution.

$$\boldsymbol{v}^{g,p} = \mathtt{D}\boldsymbol{v}^{g',p} + \mathtt{C}[r_1^{g,p}(\boldsymbol{s}^{g',p} - \boldsymbol{\omega}^{g',p}) + r_2^{g,p}(\boldsymbol{q}^{g'} - \boldsymbol{\omega}^{g',p})] \quad (8)$$

$$\boldsymbol{\omega}^{g,p} = \boldsymbol{\omega}^{g',p} + \boldsymbol{v}^{g,p} \quad (9)$$

The initial velocity $\boldsymbol{v}^{0,p}$ is set to the K-vector of zeros so that $\boldsymbol{v}^{1,p}$ equals a randomly weighted sum of the updating target's ($\boldsymbol{\omega}^{g',p}$) deviations from the two types of best-so-far candidate solutions.

Once the updated solution $\boldsymbol{\omega}^{g,p}$ has been thus computed, $\boldsymbol{s}^{g,p}$ is re-evaluated for use in the next generation: $\boldsymbol{s}^{g,p}$ equals $\boldsymbol{\omega}^{g,p}$ if $SLL(\boldsymbol{\omega}^{g,p}) > SLL(\boldsymbol{s}^{g',p})$ and $\boldsymbol{s}^{g',p}$ otherwise. Then, $\boldsymbol{q}^{g}$ is also re-evaluated and set to $\boldsymbol{s}^{g,p}$ when $SLL(\boldsymbol{s}^{g,p}) > SLL(\boldsymbol{s}^{g,p'})$ for all $p' \neq p$. In the PSO context, the terminal population of $\mathtt{P}$ candidate solutions refers to the collection of $\boldsymbol{s}^{G,p}$ for $p = 1, 2, \cdots, \mathtt{P}-1, \mathtt{P}$, instead of $\boldsymbol{\Omega}^{G}$ per se. It is the best solution in that collection, which by definition is $\boldsymbol{q}^{G}$, that is passed to a gradient-based algorithm for further improvement.

The acceleration constant $\mathtt{C}$ can be viewed as a step size parameter, much as the amplification factor $\mathtt{F}$ in the DE updating rule. The inertia weight $\mathtt{D}$ controls the tendency to continue flying in the existing direction of parametric changes. $\mathtt{C}$ is often set to 2 or less, as in the seminal study of Eberhart and Kennedy (1995). Gilli and Schumman (2010) suggest that setting $\mathtt{D}$ to a number less than 1 tends to result in better performance than setting it to 1 as in the seminal study.

4 Main case studies

This section explores the use of the DE- and PSO-assisted strategies to estimate GMNL. Each strategy passes a fully iterated DE or PSO solution as a starting point to a gradient-based algorithm to obtain the final solution. The DE- and PSO-assisted strategies are tools to improve the chance of finding the global maximum. Like any other estimation strategy, they are not guaranteed to find the global maximum. From a practitioner's standpoint, two empirical performance issues may thus be of primary interest.

The first issue is how frequently these estimation strategies can find a solution which is at least as good as the best that can be obtained using a conventional strategy. This directly relates to whether the DE- and PSO-assisted strategies are a useful addition to the practitioner's toolkit. Starting value search strategies are not part of the common reporting practice. Our own experience and conversation with colleagues, however, suggest that most practitioners would follow a similar approach as Greene and Hensher (2010, p.418) and Knox et al. (2013, p.74): the conventional strategy is to start from the estimated special cases of GMNL.

The second issue is whether some configurations of DE and PSO algorithms are conducive to finding such a solution repeatedly. This pertains to how easily the DE- and PSO-assisted strategies can be implemented in practice. As discussed earlier, each algorithm involves tuning parameters affecting how candidate solutions get updated over generations. Without knowing what these parameters need be set to, the DE- and PSO-assisted strategies would be only slightly less ambiguous than the generic advice to "try a range of starting values."

Two empirical case studies are presented below to illustrate the performance issues in detail. The data come from Pap Smear test and Pizza A choice experiments analyzed by the developers of GMNL (Fiebig et al., 2010; Keane and Wasi, 2013), and are available for download from the *Journal of Applied Econometrics* Data Archive page for Keane and Wasi (2013). Further information on these data sets is given in Fiebig et al. (2010, p.404). All estimation results have been obtained using Stata 12.1. Our Online Appendix provides a further summary of the computational settings, and can be accessed at: `https://goo.gl/Bwws7h`.

4.1 Conventional estimation strategy

Implementing the conventional estimation strategy is seemingly straightforward. It entails estimating initially a model which is nested within GMNL, and then using the results to start the GMNL estimation run. This process is to be repeated for different nested models, and the best out of several resulting GMNL solutions is picked as the preferred solution.

In practice, the conventional estimation strategy is only slightly more, if at all, straightforward than implementing the DE- and PSO-assisted strategies. Since nested models include fewer parameters, they provide estimated starting values for only some GMNL parameters; the practitioner needs to select custom starting values for the rest, and this selection may affect the final GMNL solution. The practitioner also needs to decide how the intermediate solutions are to be computed. All nested models but MNL

Table 1: Starting values based on special cases of GMNL

	MNL	MIXL	SMNL	GMNL-I	GMNL-II
$\boldsymbol{\beta}$	*Est.*	*Est.*	*Est.*	*Est.*	*Est.*
$\boldsymbol{\sigma}$	0.10	*Est.*	0.10	*Est.*	*Est.*
τ	0.25	0.25	*Est.*	*Est.*	*Est.*
γ	0	0	0	0	0

Est. indicates that the restricted model produces the relevant parameter estimates that can be directly used as starting values for GMNL.

have non-concave simulated likelihoods with potentially many maxima. Moreover, both GMNL-I and GMNL-II nest MIXL and SMNL, both of which in turn nest MNL.

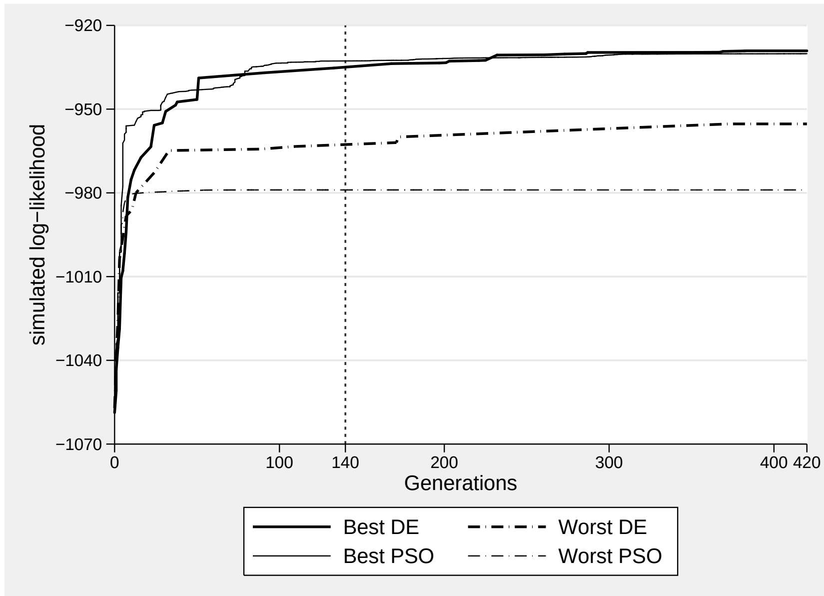
Table 1 summarizes the custom values we combined with each nested model's estimates to construct a starting point for GMNL. The MNL starting point draws on the default setting of Stata's *gmnl* command (Gu et al., 2013) and provides a basis for specifying other starting points. MIXL and SMNL were estimated from the same MNL starting point, ignoring irrelevant parameters. GMNL-I (GMNL-II) was estimated three times, once from each of the MNL, MIXL and SMNL starting points, again ignoring irrelevant parameters; GMNL, in turn, was then estimated once from each of the three potential GMNL-I (GMNL-II) starting points, though only the best of the three resulting GMNL solutions is reported below.

4.2 DE- and PSO-assisted estimation strategies

The DE and PSO algorithms require, as user inputs, the population size `P` and the number of generations `G`. In addition, both algorithms require an initial population of `P` candidate solutions that they can improve over `G` generations.

Following the common practice, we set `P`=10K where K is the number of estimated parameters. We also set `G`=10K, as preliminary experimentation with simulated data sets suggested that both algorithms tended to slow down substantially around the 10K^{th} generation. To illustrate this slowdown in an empirical context, Figure 1 plots how a selection of DE and PSO starting points used in the first case study (Pap Smear) would have varied had `G` been set to 420 (or 30K) instead of 140 (or 10K). It should be emphasised that the DE and PSO solutions at the 10K^{th} generation are used as starting points for further optimization, not as the final solutions. All of the final solutions are

Figure 1: Pap Smear: selected update paths over generations



Best DE (PSO) and Worst DE (PSO) refer to the DE (PSO) starting points that led to the best and worst final solutions in the Pap Smear case study in Section 4.3. The figure plots how these starting points had been updated until the 140th generation, at the end of which they were passed to the gradient-based algorithm for further optimization to obtain the final estimation results. The figure also plots what these starting points would have become if the algorithm continued without termination until the 420th generation.

obtained by executing the gradient-based algorithm from the DE and PSO starting points.

The initial population of `P` solutions is generated as follows. For the GMNL parameters to be estimated $\boldsymbol{\omega} = \{\boldsymbol{\beta}, \tau, \gamma, \boldsymbol{\sigma}\}$, consider the bounds given by $\boldsymbol{l} = \{\boldsymbol{b}_{MNL}, 0, 0, \mathbf{0}\}$ and $\boldsymbol{u} = \{3 \times \boldsymbol{b}_{MNL}, 2, 1, 1.5 \times \boldsymbol{b}_{MNL}\}$, where $\boldsymbol{b}_{MNL}$ is the vector of the MNL estimates and $\mathbf{0}$ is the K-vector of zeros. For each initial solution, each element of $\boldsymbol{\omega}$ is independently drawn from a uniform variable lying between the corresponding elements of $\boldsymbol{l}$ and $\boldsymbol{u}$.

The updating process of each algorithm requires two tuning parameters as additional user inputs: amplification factor `F` and cross-over probability `Cr` in case of DE, or the acceleration constant `C` and the inertia weight `D` in case of PSO. We follow Gilli and Schumann (2010) in experimenting with 16 pairs, or configurations, of those tuning parameters per algorithm: a DE configuration is in $\texttt{F} = \{0.2, 0.4, 0.6, 0.8\} \times \texttt{Cr} = \{0.2, 0.4, 0.6, 0.8\}$, while a PSO configuration is in $\texttt{C} = \{0.5, 1.0, 1.5, 2.0\} \times \texttt{D} = \{0.5, 0.75, 0.9, 1.0\}$. The resulting configurations are spaced broadly enough to provide indicative evidence for future applications on what tuning parameter values could be narrowly searched over for further fine-tuning of each algorithm.

Since the updating process is partly random, different DE or PSO starting points would result from the same configuration when different random number seeds are specified for initialization. We have obtained 48 DE starting points and 48 PSO starting points, by restarting each configuration three times from the same set of three seeds. In other words, the same set of three different initial populations has been used to obtain the three starting points associated with each configuration of each algorithm.

4.3 Results: Pap Smear

In this data set, each of 79 individuals faced 32 choice scenarios consisting of two options, namely get a Pap Smear test or not. These options are described by 6 different attributes, including the alternative-specific constant (ASC) for the get-test option. Estimating the mean ($\boldsymbol{\beta}$) and standard deviation ($\boldsymbol{\sigma}$) of the canonical random coefficient on each attribute results in 14 GMNL parameters.

Table 2 reports in descending order the simulated log-likelihood values (logL hereafter) of the solutions obtained by applying the conventional strategy, along with the usual diagnostics for checking convergence to a local optimum. Stata classifies all solutions as "converged", implying that the Hessian (H) is negative definite and the weighted gradient norm ($g'H^{-1}g$) is smaller than -1E-5 in magnitude. Further inspection suggests that only the MNL-based solution gives warning signs: the inf-norm of the gradient ($\|g\|_\infty$) deviates far way from zero and the Hessian condition number ($\kappa(H)$) exceeds one over the square root of Stata's machine precision. But this is the worst solution which is unlikely to be reported by a practitioner who tries alternative starting points.

The best conventional solution results in logL of -931.065. It is also a type of local maximum which practitioners may find particularly convincing as a candidate for the global maximum, because it can be reached from two different starting points, namely MIXL and GMNL-II. The negligible difference between their convergence diagnostics arises because the MIXL-based estimates differ marginally from the GMNL-II-based estimates, in or after the fifth decimal place.

The DE- and PSO-assisted estimation strategies find several solutions which improve on the best conventional solution. The best solution is a DE-assisted one, resulting in logL of -925.378. Table OA1 in the Online Appendix reports the logL results from all 3

Table 2: Pap Smear: conventional solutions

Starting point	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
MIXL	-931.065	8.64E-07	-2.06E-14	998.8549
GMNL-II	-931.065	2.92E-05	-4.29E-11	998.9412
SMNL	-932.133	5.46E-08	-5.30E-16	606.9426
GMNL-I	-934.091	1.95E-07	-2.12E-14	4732.774
MNL	-960.317	13.04914	-9.22E-06	1.72E+18[a]

$logL$, g and H refer to the simulated log-likelihood, its gradient (as a column vector) and Hessian respectively. The infinity norm of g, $\|g\|_\infty$, is the largest element of g in absolute value. $\kappa(H)$ is the 2-norm condition number of H, defined as $\lambda_{max}/\lambda_{min}$ where λ_{max} and λ_{min} are the largest and smallest eigenvalues of -H. Superscript a indicates that H is ill-conditioned (i.e. $\kappa(H) > 6.7\text{E}+07$).

starts of 16 configurations of each algorithm. The main features of those results may be summarized as follows. 16 of 48 DE-assisted solutions (35%) result in logL greater than -931.065, ranging from -928.034 to -925.378. Considering that some of the 48 solutions include those resulting from configurations not well-suited to the present application, a *prima facie* case exists that the DE-assisted strategy is a practically useful complement to the conventional strategy. In contrast, only 3 out of 48 PSO-assisted estimation runs (6%) result in an improved solution, ranging from -926.671 to -926.308.

Another practically attractive feature of the DE-assisted solutions is clearer indicative evidence on which configurations are likely to work well. Table 3 reports the top ten logL values found with the aid of each algorithm. A qualitative direction for fine-tuning the DE configuration to the present application would be "try a big change to the parameter estimates, but accept the resulting change only occasionally." No similar direction emerges in case of PSO, as the top ten solutions are associated with a wider range of configurations. To be specific, the top ten DE-assisted solutions are overly represented by configurations specifying a large amplification factor `F` (0.6 and 0.8) and a small cross-over probability `Cr` (0.2 and 0.4). When restricting attention to the four implied configurations, 9 out of 12 DE-assisted estimation runs (75%) find an improved solution, and 4 of those 9 runs reach the highest logL of -925.378. In contrast, a small `F` (0.2 and 0.4) appears not well suited, regardless of the accompanying `Cr`: only 2 of such 28 DE-assisted runs find an improved solution, none of them reaching the highest logL.

The highest logL has been reached from 6 different DE starting points and displays appropriate convergence diagnostics. Of course, as in the case of the best conventional

Table 3: Pap Smear: 10-best DE- and PSO-assisted solutions

A. DE-assisted solutions

F	Cr	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.8	0.6	**-925.378**	1.44E-07	-1.78E-15	2033.113
0.8	0.2	**-925.378**	9.77E-07	-1.22E-14	2033.185
0.8	0.2	**-925.378**	6.64E-05	-3.18E-11	2033.347
0.6	0.6	**-925.378**	8.07E-05	-1.35E-10	2033.3
0.6	0.4	**-925.378**	0.0001	-2.21E-10	2033.109
0.8	0.2	**-925.378**	0.000438	-3.75E-09	2033.296
0.8	0.4	**-925.409**	3.92E-07	-3.02E-15	3018.498
0.8	0.4	**-925.409**	9.32E-05	-2.26E-10	3018.577
0.6	0.2	**-926.308**	5.84E-06	-1.37E-11	936.3411
0.6	0.2	**-926.308**	0.00062	-4.03E-08	936.369

B. PSO-assisted solutions

C	D	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
1.5	0.90	**-926.308**	4.74E-06	-9.87E-13	936.3309
1.5	1.00	**-926.308**	0.000377	-4.39E-08	936.2917
2	0.90	**-926.671**	5.08E-08	-3.56E-17	1240.651
1.5	0.75	-932.176	5.49E-05	-6.57E-12	4973.183
0.5	0.90	-932.176	0.000197	-1.04E-10	4969.639
1	0.75	-932.376	7.26E-09	-6.85E-18	2174.327
2	0.50	-932.376	5.33E-08	-9.91E-16	2174.169
0.5	1.00	-932.376	4.00E-07	-1.64E-14	2173.287
1	0.90	-934.091	1.90E-08	-2.92E-16	512.7021
0.5	0.50	-934.091	1.02E-07	-6.73E-16	512.736

`F`, `Cr`, `C` and `D` indicate tuning parameter values leading to relevant starting points. *logL* is in bold if it is greater than the highest *logL* (MIXL starting point) in Table 2. See notes to Table 2 for other information.

solution, such repeatability does not imply that the underlying solution is the global maximum. Verifying that a particular solution is the global maximum is considered to be beyond the scope of our study because, as far as we are aware, no definitive guideline exists on how such verification is to be performed. We have, however, verified that the best conventional solution is not the global maximum. Our present and subsequent analysis focuses on the consequences of basing an empirical analysis on the best conventional solution when a DE- or PSO-assisted solution is capable of achieving a higher logL.

Table 4 reports parameter estimates at the second-worst and best conventional solutions, and at the best DE-assisted solution. The second-worst conventional solution

Table 4: Pap Smear: GMNL parameter estimates

	A. 2nd worst conv.		B. Best conventional		C. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
If know doctor	1.367***	(0.290)	1.202***	(0.240)	1.329***	(0.286)
	[2.764***]	(0.515)	[1.803***]	(0.246)	[2.340***]	(0.377)
If doctor is male	-3.595***	(0.657)	-2.196***	(0.339)	-2.775***	(0.556)
	[3.828***]	(0.642)	[2.760***]	(0.405)	[3.472***]	(0.479)
If test is due	5.565***	(1.211)	4.763***	(0.650)	4.969***	(0.824)
	[4.691***]	(0.911)	[3.530***]	(0.451)	[3.478***]	(0.553)
If doctor recommends	3.090***	(0.689)	1.835***	(0.293)	2.226***	(0.422)
	[2.943***]	(0.559)	[1.681***]	(0.254)	[1.201***]	(0.238)
Test cost	-0.339***	(0.101)	-0.327***	(0.094)	-0.245**	(0.096)
	[0.602***]	(0.165)	[0.022]	(0.054)	[0.180**]	(0.076)
ASC for test	-3.852***	(1.056)	-1.507***	(0.346)	-2.281***	(0.512)
	[4.140***]	(0.747)	[4.447***]	(0.517)	[4.099***]	(0.607)
γ	0.102**	(0.045)	0.081	(0.054)	0.152***	(0.055)
τ	1.304***	(0.230)	0.940***	(0.144)	0.962***	(0.158)
logL	-934.091		-931.064		-925.378	

For each named attribute, the corresponding elements of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ (in [.]) are reported. "The 2nd worst conv." and "Best conventional" respectively refer to GMNL-I and MIXL/GMNL-II starting point solutions in Table 2. "Best DE-assisted" refers to the first 6 solutions in Table 3. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

(Solution A) results from the GMNL-I starting point, and is the worst one out of conventional solution with acceptable convergence diagnostics. In terms of logL, the best conventional solution (Solution B) gains over Solution A by some 3 points, and there are marked differences between the coefficient estimates: the mean of "ASC test", in particular, is about 2.5 times larger in Solution A than in Solution B (-3.85 vs. -1.51) and many other estimates disagree even on the first significance figures.

There are less pronounced differences between the best DE-assisted solution (Solution C) and the best conventional solution (Solution B), despite that C improves on B by 6 logL points, or twice as much as B improves on A. The main difference between the solutions is that while solution B supports simplifying the model to a more parsimonious GMNL-II model with a non-random test cost coefficient, solution C does not support such a simplification as both the estimate of γ and the standard deviation of the cost coefficient are significant and non-trivial. The remaining differences are not such that it becomes immediately obvious from simple inspection whether policy-relevant

Table 5: Pap Smear: simulated WTP distributions

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
If know doctor:					
2nd worst conv.	-121	-35	8	51	145
Best conventional	-39	-2	36	76	114
Best DE-assisted	-156	-32	41	122	284
If doctor is male:					
2nd worst conv.	-244	-96	-27	45	206
Best conventional	-182	-128	-67	-8	48
Best DE-assisted	-455	-212	-83	21	207
If test is due:					
2nd worst conv.	-292	-64	41	135	340
Best conventional	-5	65	144	222	293
Best DE-assisted	-184	45	156	321	682
If doctor recommends:					
2nd worst conv.	-162	-35	21	79	205
Best conventional	-14	19	57	94	129
Best DE-assisted	-73	28	69	135	287

Figures are in $s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in Table 4. p(Q) denotes the Q^{th} percentile of the simulated distribution.

statistics derived from these solutions, such as the median willingness-to-pay (WTP) and the predicted choice probability, would be substantively different.[2]

To facilitate further comparisons, Table 5 reports selected percentiles of WTP distributions simulated from solutions A, B and C. As expected from the earlier comparison of A with B, these two solutions imply quite different median WTP, the primary statistic on which practitioners are likely to focus (e.g. Small et al., 2005). The implied WTP distributions of B and C, on the other hand, are only slightly different at the median. The main difference between those two solutions is that due to heterogeneity in the cost coefficient which is only picked up by C, the interpercentile ranges of WTP are much more pronounced for C than B. As a result, conclusions regarding the dispersion of the WTP distribution implied by B may require reconsideration.

Table 6 compares the three solutions in terms of the predicted changes in the probability of choosing the Pap Smear test in response to attribute level variations. The

[2]The WTP for a specific attribute is the utility coefficient on that attribute divided by the absolute value of the utility coefficient on the price or cost attribute. The WTP distribution can be simulated first by making simulated draws for all utility coefficients according to equation (2), and then computing relevant ratios of those simulated coefficients.

Table 6: Pap Smear: predicted choice probabilities

	A	B	C
Base choice probability	0.45	0.57	0.53
Change when test is not due	-0.24	-0.27	-0.26
Change when don't know doctor	-0.06	-0.06	-0.07
Change when doctor is female	+0.15	+0.12	+0.15
Change when doctor recommends	+0.12	+0.09	+0.11
Change when test cost is zero	+0.04	+0.05	+0.04

A, B and C are respectively based on 100,000 draws from the joint density of utility coefficients according to "2nd worst conv.", "Best conventional" and "Best DE-assisted" solutions in Table 4. The base choice probability is the probability of choosing a test (over no test) when the test is due, the patient knows the doctor, the doctor is male, the doctor makes no recommendation, and the cost is \$30. Each row reports how this probability changes when each attribute changes from its base level.

baseline specification of the attribute levels has been motivated by what Johar et al. (2013, p.1853) find plausible in the Australian context. As in the case of the median WTP, solutions B and C agree on the substantive conclusions, predicting changes of similar magnitudes and indicating that under the baseline scenario, the test is more likely to be chosen than not. In this case, however, solution A also yields almost the same results as the others, apart from that in line with its large and negative ASC, it predicts a smaller baseline probability of the test (0.45) than B (0.57) and C (0.53). This robustness may stem from the same source as the difficulties of finding the global maximum, namely that different combinations of parametric values lead to similar probabilities or likelihoods.

4.4 Results: Pizza A data

In this data set, each of 178 individuals faced 16 choice scenarios consisting of two hypothetical pizza delivery services. These services are described by 8 different attributes. Estimating the mean and standard deviation of the canonical random coefficient on each attribute results in 18 GMNL parameters.

Table 7 reports logL values attained by the conventional solutions. The MIXL and GMNL-II starting points again turn out to be two best conventional starting points. But this time, only GMNL-II leads to the highest logL of -1361.84. All conventional solutions, including the worst one, display acceptable convergence diagnostics.

Table 7: Pizza A: conventional solutions

Starting point	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
GMNL-II	-1361.84	1.45E-05	-4.88E-12	173280.2
MIXL	-1365.17	1.30E-06	-1.58E-12	20000.77
MNL	-1368.44	3.98E-05	-1.51E-09	182428.5
GMNL-I	-1374.45	0.003018	-1.71E-08	3409.606
SMNL	-1395.5	4.60E-06	-8.35E-13	442.66

See notes to Table 2.

Table 8 reports the top ten logL values attained by the DE- and PSO-assisted solutions. The full set of the DE- and PSO-assisted estimation runs are available in Table OA2 of the Online Appendix. The results agree with the Pap Smear results on two broad conclusions. First, the best solution (logL = -1356.80) is obtained by the DE-assisted strategy. Second, the DE-assisted strategy outperforms the PSO-assisted strategy in terms of finding a solution improving on the best conventional solution: 42% or 20 out of 48 DE-assisted solutions, and 23% or 11 out of 48 PSO-assisted solutions, improve on the best conventional solution.

The current results, however, are quite different from the previous results in one important dimension. 11 DE-assisted solutions (23%) and 4 PSO-assisted solutions (8%) have been declared "not converged" by Stata, because the associated Hessian is not negative definite and/or $g'H^{-1}g$ exceeds the tolerance level. Importantly, as the upper panel of Table 8 shows, the clear sign of non-convergence is present in the four best solutions that we have obtained.

Since these are symptoms of an empirically underidentified model, we followed the advice of Chiou and Walker (2007) and re-estimated the model with a higher number of simulation draws (10,000), using as starting point the best conventional solution. As Chiou and Walker point out, using a larger number of draws unmasks empirical underidentification: while the best conventional solution displays acceptable convergence diagnostics at 500 draws, the new estimation run failed to attain convergence. Thus, in the present application, the use of the DE- and PSO-assisted strategies leads to a practically different implication from the conventional strategy: namely, that the model needs to be simplified before the parameter estimates can be readily interpreted.

Table 8: Pizza A: 10-best DE- and PSO-assisted solutions

A. DE-assisted solutions

F	Cr	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.6	0.8	**-1356.8**	18479.79	-51.2587^n	-1.42E+19
0.8	0.4	**-1357.17**	1665.752	-0.16408^n	-3.35E+20
0.6	0.2	**-1357.17**	1887.993	-0.16469^n	-5.19E+20
0.6	0.4	**-1357.17**	298.4716	-0.16521^n	6360351
0.6	0.2	**-1357.53**	0.002223	-2.33E-06	4232703
0.6	0.4	**-1357.64**	0.000897	-4.58E-07	2195944
0.8	0.8	**-1357.64**	0.002647	-1.12E-06	2567936
0.4	0.8	**-1359.03**	0.000146	-4.24E-10	41664.38
0.8	0.8	**-1359.11**	4.60E-06	-5.65E-11	175508.2
0.4	0.2	**-1359.11**	0.001924	-4.17E-09	171543.3

B. PSO-assisted solutions

C	D	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
1.5	0.5	**-1359.3**	0.002958	-1.89E-07	177189.4
1.5	0.75	**-1360**	0.001075	-2.63E-06	184407.5
1	0.9	**-1360.09**	0.000029	-2.40E-08	219779.6
2	0.5	**-1360.29**	6.21E-05	-1.26E-10	32379.59
2	1	**-1360.29**	0.000188	-4.50E-10	32251.78
2	1	**-1360.29**	0.000264	-9.16E-10	32340.78
0.5	1	**-1360.71**	0.00854	-9.80E-09	218317.4
1	0.9	**-1360.76**	1.71E+12	-0.02901^n	.
1	0.5	**-1360.79**	0.000654	-1.94E-08	448585.6
2	0.75	**-1360.9**	0.000794	-1.57E-06	823405.7

logL is in bold if it is greater than the highest *logL* (GMNL-II starting point) in Table 7. Superscript n indicates that Stata has declared convergence failure since $|g'H^{-1}g|$ exceeds the tolerance criterion (1E-5). See notes to Table 2 for other informaiton.

5 Further analysis

The results described in the previous section suggest that the DE- and PSO-assisted estimation strategies can be a useful tool for improving the chance of finding the global maximum in empirical applications. Between the two strategies, the DE-assisted strategy appears to be the better choice since it improves on the conventional solution more frequently and is more consistent in terms of which configurations are likely to perform well. The best conventional and DE-assisted solutions have led to somewhat (Pap Smear) and quite (Pizza A) different substantive conclusions based on the estimated GMNL models.

The Online Appendix reports an extensive set of results from further case studies, which echo the relatively superior performance of the DE-assisted strategy. The additional case studies include applications of the DE- and PSO-assisted strategies to two larger empirical datasets (Holiday A and Mobile Phone) of Fiebig et al. (2010), as well as to simulated data sets. We also re-analyze the Pap Smear and Pizza A data sets, using alternative hybrid estimation strategies which exploit the DE and PSO algorithms jointly with the Nelder-Mead algorithm.[3] Finally, we repeat all of our four empirical case studies in the new context of estimation of the mixed logit (MIXL) model, instead of the GMNL model. The Online Appendix can be accessed at: `https://goo.gl/Bwws7h`.

6 Conclusion

In this paper, we have proposed an estimation strategy which uses the differential evolution (DE) and particle swarm optimization (PSO) algorithms to obtain starting values for random parameter logit models. Our findings suggest that the DE-assisted strategy can be a very effective tool to diagnose the adequacy of the modeling results obtained using the conventional strategy. The DE configuration ($\texttt{F} = 0.8$, $\texttt{Cr} = 0.2$) performs particularly well, and may serve as a baseline configuration in similar applications.

Our results clearly suggest that repeatedly finding a particular maximum from several starting points is not reliable evidence that it is the global maximum. Given the difficulties of verifying the global maximum in empirical work, it appears prudent to embrace the recommendation that Knittel and Metaxoglou (2014) make in a different context of non-linear optimization: namely, to report the main differences across several optima found during the estimation process.

We conclude with a few remarks on the estimation run time of the DE-assisted strategy relative to that of the conventional strategy. Comparing the run time is inherently difficult because the estimation issue of interest is not to locate a unique maximum in the fastest time but to locate the best of several possible maxima. The sensitivity of a gradient-based optimizer's run time to starting points poses another source of complication: in the Pap Smear case study, for example, conventionally estimating the GMNL model took as little as 17 minutes (from the MIXL starting point) to 11 hours (from the MNL starting point). With these caveats in mind, we note that in all of our empirical case studies, two DE-assisted estimation runs using ($\texttt{F} = 0.8$, $\texttt{Cr} = 0.2$) required a com-

[3] We thank an anonymous reviewer for this suggestion.

parable amount of time as the conventional strategy of searching over major special cases of the final model: continuing with the Pap Smear example, each DE-assisted run using this configuration took 1.5 hours whereas the conventional strategy took a combined total of 3.5 hours even when we overlook the exceptional 11-hour run from the MNL starting point. In every empirical case study and given the same configuration, at least two out of three restarts of the DE-assisted strategy located a higher maximum than the conventional strategy. These findings suggest that running two or three restarts of the DE-assisted strategy would make an effective and computationally feasible addition to the empirical practitioner's toolkit.

References

[1] Chiou, L., Walker, J. 2007. Masking identification of discrete choice models under simulation methods. *Journal of Econometrics* **141**: 683-703.

[2] Das, S., Suganthan, P. 2011. Differential evolution: a survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation* **15**: 4-31.

[3] Dorsey, R., Mayer, W. 1995. Genetic Algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features. *Journal of Business and Economic Statistics* **13**: 53-66.

[4] Eberhart, R., Kennedy, J. 1995. A new optimizer using particle swarm theory. *Proceedings of the Sixth International Symposium on Micromachine and Human Science*: pp. 39-43.

[5] Fiebig, D., Keane, M., Louviere, J., Wasi, N. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* **29**: 393-421.

[6] Gilli, M., Schumann, E. 2010. Robust regression with optimisation heuristics. In: Brabazon A, O'Neill M (Eds.), *Natural Computing in Computational Finance* Volume 3. Springer, pp.9-30.

[7] Gilli, M., Winker, P. 2009. Heuristic optimization methods in econometrics. In: Besley D, Kontoghiorghes E (Eds.), *Handbook of Computational Econometrics*. Wiley, pp.81-120.

[8] Greene, W., Hensher, D. 2010. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation* **37**: 413-428.

[9] Gu, Y., Hole, A., Knox, S. 2013. Fitting the generalized multinomial logit model in Stata. *Stata Journal* **13**: 382-397.

[10] Harding, M., Hausman, J. 2007. Using a Laplace approximation to estimate the random coefficients logit model by nonlinear least squares. *International Economic Review* **48**: 1311-1328.

[11] Huber, J., Train, K. 2001. On the similarity of Classical and Bayesian estimates of individual mean partworths. *Marketing Letters* **12**: 259-269.

[12] Johar, M., Fiebig, D., Haas, M., Viney, R. 2013. Using repeated choice experiments to evaluate the impact of policy changes on cervical screening *Applied Economics* **45**: 1845-1855.

[13] Keane, M., Wasi, N. 2013. Comparing alternative models of heterogeneity in consumer choice behavior.*Journal of Applied Econometrics* **28**: 1018-1045.

[14] Knittel, C., Metaxoglou, K. 2014. Estimation of random-coefficient demand models: two empiricists' perspective. *Review of Economics and Statistics* **96**: 34-59.

[15] Knox, S., Viney, R., Gu, Y., Hole, A., Fiebig, D., Street, D., Haas, M., Weisberg, E., Bateson, D. 2013. The effect of adverse information and positive promotion on women's preferences for prescribed contraceptive products. *Social Sciences and Medicine* **83**: 70-80.

[16] McFadden, D., Train, K. 2000. Mixed MNL models for discrete response. *Journal of Applied Econometrics* **15**: 447-470.

[17] Revelt, D., Train, K. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics* **80**: 647-657.

[18] Storn, R., Price, K. 1997. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11**: 341-359.

[19] Train, K. 2008. EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling* **1**: 40-69.

[20] Train, K. 2009. *Discrete choice methods with simulation* (2nd ed.). Cambridge University Press: New York.

Online Appendix for:

The use of heuristic optimization algorithms to facilitate maximum simulated likelihood estimation of random parameter logit models

Arne Risa Hole† and Hong Il Yoo‡

†Department of Economics, University of Sheffield
a.r.hole@sheffield.ac.uk

‡Durham University Business School, Durham University
h.i.yoo@durham.ac.uk

November, 2016

Online Appendix

This online appendix reports computational results that support the discussion presented in Section 4 and Section 5 of the main manuscript. All subsequent references to sections, and to tables without the OA prefix, correspond to the main manuscript.

4. Main case studies

The case studies take as given the preferred GMNL specifications of Fiebig et al. (2010) and Keane and Wasi (2013), and aim at estimating parameters $\boldsymbol{\beta}, \tau$, γ and $\boldsymbol{\sigma}$. Section 2 of the main manuscript provides further information on the model specification and parameters. $\boldsymbol{\sigma}$ denotes the square-root of the diagonal elements of $\boldsymbol{\Sigma}$; the off-diagonal elements are assumed to be zero. There is no observed scale heterogeneity (i.e. $\boldsymbol{z}_n = \mathbf{0}$), which means that μ_n simplifies to $\exp(\overline{\mu} + \tau v_n)$. The support of γ is the entire real line as in Keane and Wasi (2013), instead of $(0, 1)$ as in Fiebig et al. (2010). All estimation strategies have been implemented in Stata 12.1, and differ only by which starting points are supplied to the final gradient-based estimation of GMNL. Following Fiebig et al. (2010), the likelihood functions are simulated by taking 500 draws from each random parameter's postulated distribution.[1] The same 500 draws of each parameter are used for all estimation strategies to obviate the interference of simulation noise.

Gradient-based optimization tasks use the *clogit*, *mixlogit* (Hole, 2007) and *gmnl* (Gu et al., 2013) Stata commands as appropriate, following the default settings of each command unless explained otherwise; these settings include the use of Stata's implementation of the Newton-Raphson algorithm. For the DE and PSO algorithms, we coded our own programs in Stata, using the same simulated likelihood evaluator as *gmnl*.

Our conventional strategy finds solutions which are different from what Keane and Wasi (2013) report. Some of our solutions result in higher, and others worse, simulated log-likelihoods than the corresponding figures in that study. In addition to variations in the process of constructing starting points, such discrepancy may be attributed to different computing environments (Stata and Matlab), for example in terms of pseudo-random number generation. We do not pursue the exact source of the discrepancy because our case studies are not intended as replication exercises. Moreover, even within the Stata computing environment, we find a range of different solutions from different starting points.

[1] Keane and Wasi (2013) do not report the number of simulated draws used, but comparisons of their MIXL and SMNL results with Fiebig et al. (2010) suggest that it is also 500.

Table OA1. Pap Smear: all DE- and PSO-assisted solutions in Section 4

A. DE-assisted solutions

`F`	`Cr`	Start 1	Start 2	Start 3
0.2	0.2	-937.763	-940.832	-934.466
0.2	0.4	-934.466	-934.466	-935.134
0.2	0.6	-934.091	-934.6	-934.091
0.2	0.8	**-926.384**	-934.091	-940.832
0.4	0.2	-934.603	-934.091	-934.091
0.4	0.4	-934.091	-934.091	-934.814
0.4	0.6	-934.091	-934.091	-934.091
0.4	0.8	**-926.384**	-934.091	-932.376
0.6	0.2	**-926.308**	-934.091	**-926.308**
0.6	0.4	**-926.384**	-934.814	**-925.378**
0.6	0.6	**-926.384**	**-925.378**	**-926.384**
0.6	0.8	**-926.384**	-931.783	-936.769
0.8	0.2	**-925.378**	**-925.378**	**-925.378**
0.8	0.4	-935.455	**-925.409**	**-925.409**
0.8	0.6	**-928.034**	**-925.378**	-937.897
0.8	0.8	-949.114	-934.091	-937.826

B. PSO-assisted solutions

`C`	`D`	Start 1	Start 2	Start 3
0.5	0.5	-936.018	-934.091	-936.018
0.5	0.75	-934.603	-937.763	-936.153
0.5	0.9	-934.091	-932.176	-936.043
0.5	1	-934.091	-936.559	-932.376
1	0.5	-934.091	-946.345	-934.091
1	0.75	-932.376	-936.518	-934.49
1	0.9	-934.091	-934.091	-936.018
1	1	-934.091	-934.091	-936.043
1.5	0.5	-936.018	-971.979	-934.091
1.5	0.75	-932.176	-957.961	-942.731
1.5	0.9	-934.603	-934.129	**-926.308**
1.5	1	-938.321	-934.603	**-926.308**
2	0.5	-932.376	-955.963	-935.977
2	0.75	-936.018	-971.979	-959.567
2	0.9	-934.603	-954.306	**-926.671**
2	1	-938.276	-957.961	-938.898

`F`, `Cr`, `C` and `D` indicate tuning parameter values leading to relevant starting points. The simulated log-likelihood at each solution is reported, and is in boldface if it exceeds the highest *logL* (MIXL starting point) in Table 2.

Table OA2. Pizza A: all DE- and PSO-assisted solutions in Section 4

A. DE-assisted solutions

`F`	`Cr`	Start 1	Start 2	Start 3
0.2	0.2	-1363.69	-1375.04	-1368.93
0.2	0.4	-1375.04	-1371.05	-1368.93
0.2	0.6	-1366.91	-1374.07	**-1361.37**
0.2	0.8	-1377.3	-1377.17	-1379.4
0.4	0.2	**-1360.9**^{n}	-1366.91	**-1359.11**
0.4	0.4	-1364.58	-1368.29^n	-1363.42^n
0.4	0.6	-1365.68	**-1360.9**^{n}	**-1360.9**^{n}
0.4	0.8	-1364.76	**-1359.03**	-1363.5
0.6	0.2	-1363.68	**-1357.53**	**-1357.17**^{n}
0.6	0.4	**-1361.8**	**-1357.17**^{n}	**-1357.64**
0.6	0.6	**-1361.8**	**-1360.29**	-1366.91
0.6	0.8	-1362.23	**-1356.8**^{n}	-1363.44
0.8	0.2	-1362.29	**-1360.79**	**-1360.73**
0.8	0.4	**-1357.17**^{n}	-1362.29	-1365.3
0.8	0.6	-1368.43^n	-1366.46	**-1360.37**
0.8	0.8	**-1357.64**	**-1359.11**	-1367.92^n

B. PSO-assisted solutions

`C`	`D`	Start 1	Start 2	Start 3
0.5	0.5	-1375.04	-1380.17	-1369.54
0.5	0.75	-1366.36	-1378.21	-1368.93
0.5	0.9	-1374.82^n	-1366.9	-1371.21
0.5	1	**-1360.71**	-1365.36	-1372.35
1	0.5	-1363.03	-1363.48	**-1360.79**
1	0.75	-1363.48	-1388.5	-1372.08
1	0.9	**-1360.09**	**-1360.76**^{n}	-1365.84
1	1	-1370.4	-1365.61	-1377.73
1.5	0.5	-1368.78	**-1359.3**	-1363.48
1.5	0.75	-1364.83	**-1360**	-1375.46
1.5	0.9	-1382.65	-1387.44	-1367.97
1.5	1	-1369.04	-1371.21	-1367.78
2	0.5	-1381.54	-1376.35^n	**-1360.29**
2	0.75	-1373.2	-1383.7	**-1360.9**
2	0.9	-1367.69	-1374.84	-1365.83^n
2	1	**-1361.65**	**-1360.29**	**-1360.29**

`F`, `Cr`, `C` and `D` indicate tuning parameter values leading to relevant starting points. The simulated log-likelihood at each solution is reported, and is in bold-face if it exceeds the highest *logL* (GMNL-II starting point) in Table 7. Superscript n indicates that Stata has declared convergence failure since $|g'H^{-1}g|$ exceeds the tolerance criterion (1E-5).

5. Further case studies

In this section we explore the applicability of the findings in Section 4 to two larger empirical datasets (Holiday A and Mobile Phone) as well as simulated data. We also explore additional heuristic algorithms, alternative model specifications and different computational settings.

5.1 Holiday A data and Mobile Phone data

This subsection compares the performance of DE- and PSO-assisted strategies using the Holiday A and Mobile Phone data sets from Fiebig et al. (2010) and Keane and Wasi (2013). These data are on individuals' choices from hypothetical holiday packages and from hypothetical mobile phones, respectively. We have used the starting points reported in Table 1 to obtain the conventional solutions. Based on the findings from Section 4 we have focused on four DE configurations in $\texttt{F} = \{0.6, 0.8\} \times \texttt{Cr} = \{0.2, 0.4\}$ and four PSO configurations in $\texttt{C} = \{1.5, 2.0\} \times \texttt{D} = \{0.75, 0.9\}$. For each data set, we have obtained twelve DE-assisted (PSO-assisted) solutions from three restarts of each DE (PSO) configuration.

The Holiday A results are presented in the following tables. Table OA3 reports the conventional solutions, while Table OA4 reports all DE- and PSO-assisted solutions. Table OA5 reports parameter estimates at the best DE-assisted solution, and at the best and worst conventional solutions, and Table OA6 reports the percentiles of the willingness-to-pay (WTP) distributions, simulated from the utility parameter estimates presented in Table OA5.

The Mobile Phone results are presented in the following tables. Table OA7 reports the conventional solutions, while Table OA8 reports all DE- and PSO-assisted solutions. Table OA9 reports parameter estimates at the best DE-assisted solution, and at the best and worst conventional solutions, and Table OA10 reports the percentiles of the WTP distributions, simulated from the utility parameter estimates presented in Table OA9.

The DE-assisted strategy improves on the best conventional solution in 11 out of 12 restarts (92%) in Holiday A, and in all of 12 restarts in Mobile Phone (100%). Furthermore, it is interesting to note that in both data sets the ($\texttt{F} = \texttt{0.8}$, $\texttt{Cr} = \texttt{0.2}$) configuration repeatedly locates the best solution we have obtained, just like it did in the Pap Smear application. The overwhelmingly better performance of the DE-assisted strategy relative to the conventional strategy may be explained by underlying computational difficulties. The Holiday A and Mobile Phone data sets have 331 and 493 individuals, respectively, far more than the 79 and 178 individuals in the Pap Smear and Pizza A data sets. Thus, the present cases require many more person-specific likelihoods be simulated. In addition, the Mobile Phone data set requires the estimation of more than 10 extra parameters in comparison with the other data sets. With such factors adding to computational difficulties, the choice of starting values may become even more important.[2]

[2]This explanation invites the question of why the DE-assisted strategy performs worse in the Pizza A appli-

Holiday A yields qualitatively similar results to Pap Smear in the previous section. The best DE-assisted solution achieves a 22.96-point higher logL than the best conventional solution (logL = -2490.92 vs -2513.88), and this difference is much larger than the 9.3 points that the latter gains over the worst conventional solution (logL = -2523.27). Yet, in terms of the parameter estimates, the difference between the best DE-assisted and best conventional solutions is not as evident as that of the best and worst conventional solutions, apart from that the best DE-assisted solution finds much less coefficient heterogeneity for 'Airline' and more for 'Peak season'. The comparisons of simulated WTP distributions lead to the same conclusion.

In Mobile Phone, it is also the case that the best DE-assisted solution gains many more logL points over the best conventional solution than the latter gains over the worst conventional solution. The logL values of the three solutions are -3937.97, -3951.66 and -3954.89 respectively. The comparisons of the parameter estimates are less straightforward in this application as it involves many more parameters, most of which are statistically insignificant at all conventional levels. It is, nevertheless, evident that the best DE-assisted solution stands out from both the best and worst conventional solutions. Several standard deviation estimates are significant only in the best DE-assisted solution, and often larger in magnitude than the corresponding estimates in one or both of the conventional solutions. Thus, if significant standard deviations are used to gauge market segments to which particular mobile phone features may appeal, the best DE-assisted solution can lead to quite different marketing decisions than the best conventional solution.

We conclude this subsection with remarks on the PSO-assisted strategy. The findings in Section 4 suggests that the performance of various PSO configurations tends to be erratic across restarts. In the absence of clearer evidence on suitable baseline configurations, we have applied those drawn from $\mathtt{C} = \{1.5, 2.0\} \times \mathtt{D} = \{0.75, 0.9\}$ to the Holiday A and Mobile Phone data sets by restarting each of the resulting four configurations three times. The results again suggest that the DE-assisted strategy outperforms the PSO-assisted strategy: the latter improves on the best conventional solution less frequently (4 out 12 restarts in Holiday A and 7 out of 12 restarts in Mobile Phone), and the best PSO-assisted solution achieves worse logL than the best DE-assisted solution (-2507.70 in Holiday A and -3949.79 in Mobile Phone).

cation that involves more individuals and parameters. One possibility is that the result is an anomaly due to empirical underidentification of the GMNL model in the Pizza A data set.

Table OA3. Holiday A: conventional solutions

Starting point	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
MNL	-2513.88	0.000046	-8.39E-11	643.0827
GMNL-II	-2516.03	1.79E-05	-2.93E-11	1897.599
GMNL-I	-2516.16	4.37E-05	-1.09E-10	302.0431
SMNL	-2517.05	0.000523	-2.67E-09	4138.69
MIXL	-2523.27	8.08E-06	-3.17E-12	9313.463

$logL$, g and H refer to the simulated log-likelihood, its gradient (as a column vector) and Hessian respectively. The infinity norm of g, $\|g\|_\infty$, is the largest element of g in absolute value. $\kappa(H)$ is the 2-norm condition number of H, defined as $\lambda_{max}/\lambda_{min}$ where λ_{max} and λ_{min} are the largest and smallest eigenvalues of -H.

Table OA4. Holiday A: all 12 DE- and PSO-assisted solutions

A. DE-assisted solutions

F	Cr	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.8	0.2	**-2490.917**	0.00049	-3.54E-08	2362.534
0.8	0.2	**-2491.486**	0.000646	-1.49E-08	604.254
0.8	0.4	**-2507.704**	0.001517	-1.58E-08	758.8516
0.6	0.4	**-2508.997**	2.24E-07	-1.65E-15	271.6472
0.6	0.4	**-2508.997**	3.88E-05	-4.89E-12	271.6682
0.6	0.4	**-2508.997**	4.01E-05	-3.50E-11	271.6631
0.8	0.2	**-2508.997**	0.000279	-1.38E-09	271.6736
0.8	0.4	**-2510.231**	6.29E-06	-1.75E-12	762.5714
0.6	0.2	**-2510.397**	2.43E-07	-2.56E-15	297.3875
0.6	0.2	**-2510.397**	2.77E-06	-2.72E-13	297.3904
0.6	0.2	**-2510.397**	0.000496	-4.76E-09	297.3281
0.8	0.4	-2513.880	1.05E-05	-2.15E-12	642.9105

B. PSO-assisted solutions

C	D	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
1.5	0.75	**-2507.704**	2.71E-06	-5.35E-14	758.9773
2	0.75	**-2508.785**	0.000994	-1.31E-07	296.8386
1.5	0.9	**-2511.953**	0.001607	-7.26E-08	3111.238
2	0.9	**-2513.807**	4.27E-07	-1.19E-14	1315.024
1.5	0.75	-2513.880	1.06E-05	-2.74E-12	643.0941
2	0.75	-2513.880	0.000017	-7.63E-12	642.8463
2	0.9	-2513.880	0.000157	-4.13E-10	642.8153
1.5	0.9	-2513.965	0.00044	-3.95E-09	527.7799
1.5	0.9	-2517.047	8.95E-06	-7.48E-13	4143.885
2	0.75	-2519.290	0.000611	-1.58E-08	1559.07
2	0.9	-2521.649	0.000237	-5.47E-10	200.2613
1.5	0.75	-2524.128	2.56E-06	-7.19E-13	5860.282

F, Cr, C and D indicate tuning parameter values leading to relevant starting points. $logL$ is in bold if it is greater than the highest $logL$ (MNL starting point) in Table OA3. See notes to Table OA3 for other information.

Table OA5. Holiday A: GMNL parameter estimates

	A. Worst conventional		B. Best conventional		C. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Price	-1.009***	(0.287)	-0.826***	(0.148)	-0.735***	(0.178)
	[0.565***]	(0.159)	[0.921***]	(0.178)	[1.000***]	(0.295)
Overseas destination	0.775***	(0.265)	0.391***	(0.094)	0.411***	(0.088)
	[4.007***]	(1.089)	[3.105***]	(0.602)	[2.922***]	(0.704)
Airline	-0.125	(0.082)	-0.083	(0.070)	-0.062	(0.079)
	[0.183*]	(0.099)	[0.552***]	(0.151)	[0.101**]	(0.047)
Length of stay	1.805***	(0.477)	1.268***	(0.261)	1.354***	(0.320)
	[1.546***]	(0.424)	[1.197***]	(0.233)	[1.306***]	(0.328)
Meal inclusion	1.796***	(0.474)	1.314***	(0.233)	1.478***	(0.403)
	[1.617***]	(0.474)	[1.130***]	(0.217)	[1.564***]	(0.354)
Local tours availability	0.722***	(0.238)	0.552***	(0.128)	0.606***	(0.190)
	[0.636***]	(0.216)	[0.529***]	(0.130)	[0.658***]	(0.178)
Peak season	0.277**	(0.114)	0.143*	(0.073)	0.146**	(0.065)
	[0.913***]	(0.298)	[0.047]	(0.077)	[0.241**]	(0.103)
4-star accommodation	2.817***	(0.763)	2.037***	(0.364)	2.023***	(0.444)
	[2.341***]	(0.627)	[2.061***]	(0.424)	[1.883***]	(0.444)
γ	-0.056**	(0.025)	-0.142***	(0.045)	-0.144***	(0.049)
τ	1.416***	(0.170)	1.205***	(0.132)	1.264***	(0.166)
$logL$	-2523.271		-2513.880		-2490.917	

The worst and best conventional solutions result from MIXL and MNL starting points respectively, and display acceptable convergence diagnostics. The best DE-assisted solution results from configuration $\texttt{F} = 0.8$ and $\texttt{Cr} = 0.2$. Table OA3 and Table OA4 provide related computational results. For each named attribute, the corresponding elements of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ (in [.]) are reported. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table OA6. Holiday A: simulated WTP distributions

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
Overseas destination:					
Worst conventional	-854	-285	151	598	1282
Best conventional	-1015	-302	68	417	1128
Best DE-assisted	-1043	-291	79	420	1170
Airline:					
Worst conventional	-87	-47	-24	-3	19
Best conventional	-217	-79	-15	52	177
Best DE-assisted	-58	-25	-12	4	36
Length of stay:					
Worst conventional	-17	168	350	573	984
Best conventional	-372	47	228	421	867
Best DE-assisted	-556	23	248	476	1014
Meal inclusion:					
Worst conventional	-36	156	346	579	1027
Best conventional	-381	65	242	424	872
Best DE-assisted	-640	0	268	521	1097
Local tours availability:					
Worst conventional	-10	66	139	234	402
Best conventional	-156	20	102	185	363
Best DE-assisted	-243	4	112	218	479
Peak season:					
Worst conventional	-181	-48	54	162	343
Best conventional	-43	15	27	43	85
Best DE-assisted	-87	-9	27	60	139
4-star accommodation:					
Worst conventional	-8	265	547	896	1534
Best conventional	-625	69	372	677	1391
Best DE-assisted	-828	36	369	700	1528

Figures are in $s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in Table 9 of the main text. p(Q) denotes the Q^{th} percentile of the simulated distribution.

Table OA7. Mobile Phone: conventional solutions

Starting point	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
GMNL-I	-3951.66	8.61E-06	-1.19E-12	131.5918
MNL	-3951.74	0.005431	-1.10E-07	1682.623
MIXL	-3952.17	0.000025	-2.15E-11	89.09985
GMNL-II	-3953.72	0.000277	-3.35E-10	151.0981
SMNL	-3954.89	0.00242	-2.05E-07	447.8617

See notes to Table OA3.

Table OA8. Mobile phone: all 12 DE- and PSO-assisted solutions

A. DE-assisted solutions

F	Cr	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.6	0.4	**-3937.969**	0.001103	-9.24E-09	3702.611
0.8	0.2	**-3937.969**	0.001453	-1.14E-08	3698.788
0.6	0.4	**-3937.969**	0.008783	-6.53E-08	3703.418
0.8	0.2	**-3937.969**	0.003234	-2.32E-07	3705.606
0.6	0.2	**-3937.969**	0.021431	-4.97E-07	3696.37
0.8	0.4	**-3938.663**	0.046538	-7.63E-07	25148.27
0.6	0.2	**-3939.908**	1.81E-05	-3.32E-12	2991.533
0.6	0.2	**-3939.908**	2.95E-05	-8.14E-12	2994.464
0.8	0.2	**-3941.913**	6.17E-05	-3.59E-09	2589.266
0.8	0.4	**-3944.020**	0.002464	-8.89E-08	20553.94
0.8	0.4	**-3944.431**	9.12E-05	-2.40E-10	12952.91
0.6	0.4	**-3945.929**	3.74E-05	-5.52E-11	1036.657

B. PSO-assisted solutions

C	D	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
2	0.9	**-3949.788**	1.29E-06	-2.00E-14	161.9144
1.5	0.9	**-3949.788**	4.89E-05	-6.30E-11	161.896
1.5	0.75	**-3949.788**	0.000218	-5.94E-10	161.9118
1.5	0.9	**-3949.788**	0.000335	-2.06E-09	161.9023
2	0.9	**-3950.458**	0.000108	-1.28E-11	862.297
1.5	0.9	**-3950.458**	0.000164	-2.52E-10	862.1408
1.5	0.75	**-3950.752**	3.94E-05	-1.23E-10	333.8932
2	0.9	-3951.756	1.48E-06	-7.81E-14	117.0487
2	0.75	-3951.756	0.000133	-2.33E-10	117.0486
2	0.75	-3951.756	0.000569	-4.41E-09	117.0475
1.5	0.75	-3952.300	0.000548	-3.50E-09	136.2395
2	0.75	-3952.829	1.65E-05	-3.01E-12	121.7244

logL is in bold if it is greater than the highest *logL* (GMNL-I starting point) in Table OA7. See notes to OA4 for other information.

Table OA9. Mobile Phone: GMNL parameter estimates

	A. Worst conventional		B. Best conventional		C. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
No voice comm.	0.045	(0.057)	0.063	(0.055)	0.064	(0.087)
	[0.134]	(0.105)	[0.069]	(0.118)	[0.037]	(0.090)
Voice dialing	0.100*	(0.054)	0.085	(0.058)	0.125	(0.085)
	[0.013]	(0.087)	[0.131]	(0.163)	[0.233***]	(0.067)
Voice operation	-0.155**	(0.063)	-0.136**	(0.060)	-0.099	(0.087)
	[0.167]	(0.103)	[0.098]	(0.143)	[0.254***]	(0.086)
No push to com.	0.056	(0.059)	0.054	(0.058)	0.043	(0.075)
	[0.161**]	(0.081)	[0.005]	(0.083)	[0.213***]	(0.077)
Push to talk	0.059	(0.056)	0.039	(0.060)	0.060	(0.075)
	[0.069]	(0.087)	[0.245***]	(0.080)	[0.206***]	(0.072)
Push to share pics/video	-0.025	(0.061)	-0.021	(0.055)	0.055	(0.074)
	[0.041]	(0.110)	[0.027]	(0.127)	[0.096*]	(0.056)
Personal e-mail	-0.035	(0.071)	-0.059	(0.057)	0.004	(0.082)
	[0.034]	(0.081)	[0.032]	(0.089)	[0.089]	(0.086)
Corporate e-mail	0.080	(0.057)	0.065	(0.056)	0.051	(0.076)
	[0.147]	(0.102)	[0.106]	(0.122)	[0.058]	(0.048)
Both e-mails	-0.060	(0.059)	-0.058	(0.057)	-0.233**	(0.094)
	[0.147*]	(0.087)	[0.085]	(0.085)	[0.085]	(0.054)
WiFi	-0.016	(0.031)	-0.023	(0.031)	-0.059	(0.045)
	[0.006]	(0.051)	[0.012]	(0.051)	[0.008]	(0.044)
USB calbe/cradle	0.095**	(0.043)	0.086**	(0.034)	0.184***	(0.061)
	[0.088]	(0.163)	[0.047]	(0.072)	[0.131***]	(0.048)
Thermometer	0.049	(0.034)	0.063*	(0.037)	0.052	(0.047)
	[0.134]	(0.083)	[0.185***]	(0.066)	[0.151***]	(0.049)
Flashlight	0.063*	(0.034)	0.045	(0.033)	0.083	(0.061)
	[0.029]	(0.061)	[0.075]	(0.069)	[0.003]	(0.062)
Price/100	-1.214***	(0.264)	-1.110***	(0.143)	-1.880***	(0.302)
	[1.096***]	(0.176)	[1.066***]	(0.144)	[0.945***]	(0.210)
ASC for purchase	-0.574***	(0.151)	-0.661***	(0.175)	-1.182***	(0.253)
	[2.542***]	(0.437)	[2.639***]	(0.281)	[4.122***]	(0.907)
γ	-0.108	(0.186)	-0.234**	(0.114)	-0.502***	(0.130)
τ	0.852***	(0.263)	-0.804***	(0.115)	1.715***	(0.162)
logL	-3954.893		-3951.662		-3937.969	

The worst and best conventional solutions result from SMNL and GMNL-I starting points respectively, and display acceptable convergence diagnostics. The best DE-assisted solution results from configurations (`F`,`Cr`) = (0.6,0.2), (0.6,0.4) and (0.8,0.2). Table OA7 and Table OA8 provide related computational results. For each named attribute, the corresponding elements of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ (in [.]) are reported. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table OA10. Mobile Phone: simulated WTP distributions

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
No voice comm.:					
Worst conventional	-16	-4	3	10	24
Best conventional	-6	1	4	9	19
Best DE-assisted	-3	1	3	5	9
Voice dialing:					
Worst conventional	-8	4	7	11	24
Best conventional	-13	-1	6	13	30
Best DE-assisted	-24	-3	6	14	35
Voice operation:					
Worst conventional	-44	-21	-9	-1	15
Best conventional	-36	-18	-10	-4	9
Best DE-assisted	-37	-14	-5	5	26
No push to com.:					
Worst conventional	-18	-5	3	12	30
Best conventional	-5	3	4	7	13
Best DE-assisted	-25	-6	2	10	27
Push to talk:					
Worst conventional	-7	-0	4	8	18
Best conventional	-32	-9	3	15	40
Best DE-assisted	-22	-5	3	10	28
Push to share pics/video:					
Worst conventional	-9	-4	-2	1	4
Best conventional	-7	-3	-2	-0	3
Best DE-assisted	-10	-1	3	6	14
Personal e-mail:					
Worst conventional	-10	-5	-2	-0	3
Best conventional	-15	-7	-4	-2	4
Best DE-assisted	-11	-3	0	4	11
Corporate e-mail:					
Worst conventional	-15	-2	5	14	31
Best conventional	-11	-1	5	10	24
Best DE-assisted	-5	-0	2	5	10

(*continued on the next page*)

Table OA10. Mobile Phone: simulated WTP distributions

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
		(continued from the previous page)			
Both e-mails:					
Worst conventional	-29	-12	-4	3	16
Best conventional	-19	-9	-4	0	9
Best DE-assisted	-27	-15	-11	-5	6
WiFi:					
Worst conventional	-4	-2	-1	-1	1
Best conventional	-6	-3	-2	-1	1
Best DE-assisted	-6	-4	-3	-2	1
USB calbe/cradle:					
Worst conventional	-8	1	6	12	26
Best conventional	-5	3	6	11	22
Best DE-assisted	-11	2	9	14	27
Thermometer:					
Worst conventional	-15	-4	3	11	25
Best conventional	-20	-4	5	15	34
Best DE-assisted	-16	-3	3	8	20
Flashlight:					
Worst conventional	-4	2	4	8	16
Best conventional	-7	-1	3	7	17
Best DE-assisted	-1	2	4	5	8

Figures are in \$s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in Table 10 of the main text. p(Q) denotes the Q^{th} percentile of the simulated distribution.

5.2 Pap Smear and Pizza A: 20 starts

Our findings so far have suggested that good baseline configurations of the DE algorithm can be drawn from $\texttt{F} = \{0.6, 0.8\} \times \texttt{Cr} = \{0.2, 0.4\}$. As explained earlier, the starting point for the algorithm is randomly determined. This sub-section explores the robustness of the configurations by restarting each of the four configurations using twenty different random number seeds, whereas the preceding analysis used three seeds. We focus on the Pap Smear and Pizza A data sets whose smaller sizes make them more amenable to a large number of estimation runs.

In each data set, the results over 80 restarts confirm that the performance of these configurations is consistently good. In the Pap Smear data, 49 out of 80 restarts (61.25%) improve on the best conventional solution. The frequency is smaller than the 75% (over 12 comparable restarts) found earlier, but still covers the majority of cases. In the Pizza A data 62 out of 80 restarts (77.5%) improve on the best conventional solution, that is with a higher frequency than the 67% found earlier.

Table OA11 reports the ten best DE-assisted solutions found from the 80 restarts in each data set. The results for the Pap Smear data suggest that (as in the case of the best conventional solution) repeatedly finding a particular maximum is not a reliable sign that it is the global maximum. Now, there are two new maxima at the logL values of -924.359 and -924.788, both of which are higher than the logL of -925.378 in the best DE-assisted solution found in Section 4, which was reached four times out of the 12 restarts from the configurations under consideration. Table OA12 reports the parameter estimates at our earlier best solution and the two new best solutions. An interesting aspect of the parameter estimates at -924.359 is that like the best conventional solution, the standard deviation of the cost coefficient is small and insignificant, in contrast with the "best DE-assisted" solution of the previous section where it is significant. Given the difficulties of verifying the global maximum in empirical work, it appears prudent to report all main differences across several maxima found in estimation runs, as Knittel and Metaxoglou (2014) recommend in the context of the Berry-Levinsohn-Pakes method of demand estimation.

While some of the new DE-assisted solutions for Pizza A also improve on our earlier best, we do not report the associated parameter estimates because there is evidence that GMNL is an empirically underidentified model for this data set: see Section 4.4.

Table OA11. 10-best DE-assisted solutions over 20 starts per configuration

A. Pap Smear

F	Cr	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.8	0.2	**-924.359**	1.07E-06	-5.45E-15	3741.156
0.8	0.2	**-924.359**	6.55E-07	-9.14E-15	3741.275
0.8	0.2	**-924.359**	2.99E-06	-7.60E-14	3741.364
0.8	0.4	**-924.359**	6.90E-06	-1.35E-13	3740.987
0.8	0.4	**-924.359**	1.27E-05	-3.03E-13	3741.41
0.8	0.2	**-924.359**	2.23E-05	-8.58E-11	3741.099
0.8	0.4	**-924.359**	0.000052	-2.18E-10	3740.992
0.8	0.4	**-924.359**	0.000286	-6.84E-10	3740.659
0.8	0.4	**-924.788**	7.61E-07	-1.34E-14	4324.144
0.8	0.2	**-924.788**	0.000341	-3.07E-10	4322.18

B. Pizza A

F	Cr	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.8	0.4	**-1352.91**	2.39E-04	-7.91E-09	283510.5
0.8	0.2	**-1353.91**	4.01E+02	-1.26E-01^n	1.57E+07
0.8	0.2	**-1354.58**	4.06E-04	-1.44E-08	124508.8
0.8	0.2	**-1355.67**	10.47107	-5.07E-02^n	433304.1
0.8	0.4	**-1355.9**	0.001573	-2.21E-08	601560.5
0.8	0.2	**-1356.84**	7.441122	-6.84E-04^n	-1.31E+19
0.6	0.4	**-1357.17**	12815.33	-4.34E-01^n	-1.17E+17
0.6	0.4	**-1357.17**	1.50E+03	-1.65E-01^n	-7.76E+20
0.6	0.2	**-1357.17**	353.063	-1.65E-01^n	8613358
0.6	0.2	**-1357.17**	503.7171	-0.16532^n	1.82E+07

Information for panel A is the same as in notes to Table 3 of the main manuscript. Information for panel B is the same as in notes to Table 7 of the main manuscript. The results in those tables have been obtained by restarting each configuration 3 times. The results in this table have been obtained by restarting each configuration 20 times.

Table OA12. Pap Smear: GMNL parameter estimates (from 20 starts per configuration)

	A. Third		B. Second		C. First	
	Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
If know doctor	1.329***	(0.286)	1.210***	(0.257)	1.179***	(0.290)
	[2.340***]	(0.377)	[1.768***]	(0.282)	[1.726***]	(0.258)
If doctor is male	-2.775***	(0.556)	-2.189***	(0.455)	-1.813***	(0.349)
	[3.472***]	(0.479)	[4.293***]	(0.636)	[3.845***]	(0.457)
If test is due	4.969***	(0.824)	5.884***	(0.939)	5.823***	(0.822)
	[3.478***]	(0.553)	[3.375***]	(0.522)	[4.703***]	(0.705)
If doctor recommends	2.226***	(0.422)	2.150***	(0.448)	2.450***	(0.405)
	[1.201***]	(0.238)	[2.022***]	(0.316)	[1.792***]	(0.303)
Test cost	-0.245**	(0.096)	-0.314***	(0.115)	-0.329***	(0.097)
	[0.180**]	(0.076)	[0.262***]	(0.091)	[0.047]	(0.056)
ASC for test	-2.281***	(0.512)	-1.687***	(0.486)	-1.104***	(0.412)
	[4.099***]	(0.607)	[4.701***]	(0.679)	[3.977***]	(0.494)
γ	0.152***	(0.055)	0.096	(0.059)	0.110***	(0.037)
τ	0.962***	(0.158)	1.076***	(0.169)	0.960***	(0.132)
logL	-925.378		-924.788		-924.359	

For each named attribute, the corresponding elements of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ (in [.]) are reported. All three solutions are DE-assisted solutions. The "Third" best solution coincides with the "Best DE-assisted" solution in Table 4 of the main text. The "Second" and "First" best solutions are from panel A in Table OA11. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

5.3 Simulated data sets: Monte Carlo evidence

In all four empirical data sets, the DE-assisted strategy has located higher maxima than both the conventional and PSO-assisted strategies. DE configurations in $\texttt{F}=\{0.6, 0.8\}\times\texttt{Cr}=\{0.2, 0.4\}$ appear viable as baseline settings for the tuning parameters in empirical work. In particular, configuration ($\texttt{F}=0.8$, $\texttt{Cr}=0.2$) has located the best solution in all data sets except Pizza A, wherein the model showed symptoms of empirical underidentification.

To check the robustness of these findings, we have carried out two experiments involving simulated data sets. Each data set is identical to the Pap Smear data set, except that simulated choices have replaced actual choices. The data generating process for the simulated choices is the GMNL model of the Pap Smear case study in Section 4.3, and the true parameter values are close to the best DE-assisted solution in Table 4.[3] In both experiments, the model's parameters are estimated by applying different strategies to each simulated data set. The computational demands of maximum simulated likelihood estimation make it practically difficult to analyze a large number of data sets using the full range of estimation strategies that we have explored so far, and for this reason each experiment focuses on a subset of the strategies selected to revisit a specific aspect of our empirical findings.[4]

The first experiment examines the extent to which the better performance of the DE-assisted strategy, especially in combination with the candidate baseline configurations, is repeated across 100 simulated data sets. For each data set, we obtain 3 conventional solutions using each of MNL, SMNL and MIXL starting points; 16 DE-assisted solutions by starting the DE algorithm once from each configuration in $\texttt{F}=\{0.2, 0.4, 0.6, 0.8\}\times\texttt{Cr}=\{0.2, 0.4, 0.6, 0.8\}$; and 16 PSO-assisted solutions by starting the PSO algorithm once from each configuration in $\texttt{C}=\{0.5, 1.0, 1.5, 2.0\}\times\texttt{D}=\{0.5, 0.75, 0.9, 1.0\}$. As before, the best solution in each data set refers to one(s) that resulted in the highest of up to 35 distinct local maxima.

The results support the earlier findings on the performance of the DE-assisted strategy and the candidate baseline configurations. The DE-assisted strategy found the best solution in 87 out of 100 data sets, whereas the PSO-assisted strategy and the conventional strategy found it in 12 and 7 data sets respectively.[5] Even when one focuses on four DE-assisted solutions using

[3] A fuller summary of the data generating process is as follows. In each data set, each agent n's utilities are simulated by making new draws of random preference parameters $\boldsymbol{\beta}_n$ and type I extreme value errors, and combining them with the actual attributes according to equation 1. Then, the utility-maximizing alternative in a choice scenario becomes the simulated choice in that scenario. The population density of $\boldsymbol{\beta}_n$ is that of the GMNL specification for the Pap Smear application in Section 4.3, and the density's parameters take values close to the best DE-assisted solution in Table 4: see Table OA14 for further information on the true parameter values.

[4] We note that the number of simulated data sets in each of our case studies (100 in one, and 1000 in the other) is relatively large for this type of estimation problem. Fiebig et al. (2010), for example, carried out a Monte Carlo study based on the Pap Smear data set using 25 data sets, and their study did not involve comparisons of solutions resulting from alternative starting points.

[5] The numbers do not add up to 100 because in six data sets, the best solution was found by multiple estimation strategies.

configurations in $\texttt{F}=\{0.6, 0.8\}\times\texttt{Cr}=\{0.2, 0.4\}$, the DE-assisted strategy found the best solution in 53 data sets, and improved on the best of the three conventional solutions in 86 data sets. Table OA13 reports the number of data sets in which each of 3 conventional, 16 DE and 16 PSO starting points led to the best solution. Like our empirical findings in section 4.3, this breakdown suggests that the MIXL starting point tends to outperform the other two conventional starting points and a good baseline DE configuration would have $\texttt{F}=\{0.6, 0.8\}$, whereas no particular PSO configuration stands out from the rest. Moreover, ($\texttt{F}=0.8$, $\texttt{Cr}=0.2$) again appears to be a particularly good choice as the baseline DE configuration for empirical work: it led to the best solution more often (in 21 data sets) than all but one configuration.

Table OA13. A summary of results across 100 simulated data sets

A. DE-assisted				B. PSO-assisted			
F	Cr	$Freq_{best}$	$logL$	C	D	$Freq_{best}$	$logL$
0.2	0.2	3	-938.529	0.5	0.5	1	-935.256
0.2	0.4	1	-938.649	0.5	0.75	1	-934.452
0.2	0.6	1	-937.874	0.5	0.9	2	-933.856
0.2	0.8	1	-940.049	0.5	1	3	-935.057
0.4	0.2	3	-932.390	1	0.5	2	-934.341
0.4	0.4	4	-932.971	1	0.75	1	-934.124
0.4	0.6	5	-932.065	1	0.9	1	-934.970
0.4	0.8	13	-929.837	1	1	1	-935.394
0.6	0.2	13	-929.048	1.5	0.5	0	-934.316
0.6	0.4	13	-928.426	1.5	0.75	0	-934.715
0.6	0.6	23	-927.441	1.5	0.9	2	-935.578
0.6	0.8	16	-928.131	1.5	1	2	-935.659
0.8	0.2	21	-927.602	2	0.5	2	-935.883
0.8	0.4	18	-928.478	2	0.75	1	-936.221
0.8	0.6	13	-929.956	2	0.9	3	-935.845
0.8	0.8	4	-931.923	2	1	1	-936.490

Each row of panel A (panel B) summarizes the results obtained by setting the tuning parameters to the values shown in the first two columns of that panel. $Freq_{best}$ is the number of data sets in which a particular strategy has located the best solution, and $logL$ is the average log-likelihood of 100 solutions. $Freq_{best}$ ($logL$) associated with the conventional strategy is 1 (-943.868) when using the MNL starting point; 5 (-932.712) when using the MIXL starting point; and 1 (-934.908) when using the SMNL starting point.

The second experiment examines whether a small number of restarts from ($\texttt{F}=0.8$, $\texttt{Cr}=0.2$) is able to help detecting the potential inadequacy of a conventional solution in 1,000 simulated data sets. For each data set, we obtain one conventional solution and one DE-assisted solution. The conventional solution uses the MIXL starting point which led to the best conventional solution in the empirical Pap Smear data set, as well as in 57 of 100 data sets in the first experiment. The DE-assisted solution uses one start from ($\texttt{F}=0.8$, $\texttt{Cr}=0.2$).

In 840 out of the 1,000 data sets, the DE-assisted strategy found a higher maximum than the conventional strategy. On average, the DE-assisted solutions achieved a 5-point increase in log-likelihood, which is similar to the log-likelihood difference between the best DE-assisted and best conventional solutions seen earlier in Table 4. As in the empirical case study, however, this difference does not seem to translate into practically different conclusions. Table OA14 reports the bias and empirical standard deviations of GMNL parameter estimates across the 1000 simulated data sets. All parameter estimates display rather limited amounts of bias, except parameter γ.[6]

Table OA14. GMNL parameter estimates: results across 1000 simulated data sets

	True	Bias		Std. Dev.	
		Conv.	DE-asst.	Conv.	DE-asst.
Mean: If know doctor	1.50	0.00	0.01	0.83	0.81
Mean: If doctor is male	-2.75	0.02	-0.08	1.22	1.17
Mean: If test is due	5.00	-0.14	0.10	1.76	1.91
Mean: If doctor recommends	2.25	0.02	0.07	0.79	0.82
Mean: Test cost	-0.25	0.01	0.00	0.15	0.15
Mean: ASC for test	-2.25	-0.16	-0.07	1.27	1.23
SD: If know doctor	2.50	0.08	0.05	0.91	0.90
SD: If doctor is male	3.50	0.27	0.10	1.41	1.33
SD: If test is due	3.50	0.31	-0.02	1.46	1.31
SD: If doctor recommends	1.25	-0.04	-0.15	0.68	0.70
SD: Test cost	0.20	0.07	0.04	0.20	0.19
SD: ASC for test	4.00	0.14	0.11	1.45	1.39
τ	1.00	-0.06	0.07	0.34	0.34
γ	0.15	8.79	7.20	493.81	323.86

Conv. summarizes 1000 conventional solutions (average log-likelihood = -932.301) using the MIXL starting point. DE-asst. summarizes 1000 DE-assisted solutions (average log-likelihood = -927.176) using configuration (`F` = 0.8, `C` = 0.2). Mean (SD) denotes the population mean (standard deviation) of a normally distributed random coefficient on a particular attribute. True reports the true parameter values used in the data generating process. Bias is the difference between the average of 1000 parameter estimates and the underlying true value. Std. Dev. is the empirical standard deviation of 1000 parameter estimates.

5.4 Pap Smear and Pizza A: Comparisons with other hybrid estimation strategies

Our DE- and PSO-assisted estimation strategies follow the tradition of other hybrid estimation strategies in econometrics (Bhat, 1997; Dorsey and Mayer, 1995) in that each strategy passes a gradient-free optimization algorithm's solution as starting point to a gradient-based algorithm.

[6] Unlike the GMNL model specification of Keane and Wasi (2013) that we have estimated, the specification of Fiebig et al. (2010) constrains γ to the (0,1) interval a priori. We note that the bias pertaining to γ remains noticeable even when the results are summarized over only those solutions (750 conventional and 841 DE-assisted) wherein the estimated γ lies within this interval.

In other disciplines, hybrid optimization strategies that combine gradient-free algorithms (Liu and Yang, 2012; Luchi and Krohling, 2015) have also attracted attention.[7] Luchi and Krohling (2015), for example, propose a hybrid strategy that involves passing a DE solution as starting point to the Nelder-Mead (NM) algorithm, and show its effectiveness in the context of non-linear integer optimization. In several econometric software packages including Stata, their "DE+NM" strategy can be readily implemented once our DE-assisted strategy has been programmed since the NM algorithm is often available as an option of a package's built-in optimizer. The DE+NM strategy may also be readily adapted as a PSO+NM strategy, by replacing a DE starting point with a PSO starting point.

This subsection reports the results from applying two alternative hybrid estimation strategies (DE+NM and PSO+NM) to each of Pap Smear and Pizza A data sets. For each data set, the DE+NM (PSO+NM) strategy used exactly the same set of 48 DE (PSO) starting points as our DE-assisted (PSO-assisted) strategy. The NM algorithm was executed under Stata's default settings by using Pfeffer's method (Baudin, 2010, pp.19-20) to construct the initial simplex. Table OA15 reports the full set of 48 DE+NM and 48 PSO+NM solutions for Pap Smear, and Table OA16 does so for Pizza A. For the corresponding DE-assisted and PSO-assisted solutions, see Table OA1 and Table OA2.

The results suggest that neither of these two alternative hybrid strategies appears well-suited to the task of estimating a random parameter logit model. The best conventional solution achieves the simulated log-likelihood of -931.065 in Pap Smear (see Table 2) and -1361.84 in Pizza A (see Table 7). In each data set, none of 48 DE+NM solutions and 48 PSO+NM solutions improves on the best conventional solution, implying that they fail to improve on the best DE-assisted solution too. Furthermore, in a pairwise comparison between a DE+NM (PSO+NM) solution and a DE-assisted (PSO-assisted) solution that share the same DE (PSO) starting point, the DE+NM (PSO+NM) solution achieves the higher simulated log-likelihood only in one out of 96 cases across the two data sets. In addition, Stata classifies none of the DE+NM and PSO+NM solutions as "converged", meaning that their Hessian (H) is not negative definite and/or their weighted gradient norm ($g'H^{-1}g$) exceeds -1E-5 in magnitude.

[7] We thank an anonymous reviewer for alerting us to the relevant literature.

Table OA15. Pap Smear: all DE+NM and PSO+NM solutions

A. DE+NM solutions

F	Cr	Start 1	Start 2	Start 3
0.2	0.2	-942.714	-950.870	-938.642
0.2	0.4	-938.317	-938.297	-949.363
0.2	0.6	-939.619	-941.781	-939.020
0.2	0.8	-943.397	-938.339	-942.384
0.4	0.2	-935.761	-935.280	-935.356
0.4	0.4	-934.322	-934.804	-934.842
0.4	0.6	-934.403	-934.403	-934.239
0.4	0.8	-926.655	-934.180	-933.670
0.6	0.2	-926.515	-934.333	-927.309
0.6	0.4	-926.874	-935.095	-927.521
0.6	0.6	-928.368	-926.477	-930.887
0.6	0.8	-927.457	-934.404	-938.219
0.8	0.2	-929.938	-927.312	-928.376
0.8	0.4	-937.279	-928.460	-927.512
0.8	0.6	-935.072	-932.998	-942.737
0.8	0.8	-949.830	-938.361	-943.642

B. PSO+NM solutions

C	D	Start 1	Start 2	Start 3
0.5	0.5	-958.292	-952.565	-957.391
0.5	0.75	-938.914	-941.970	-938.878
0.5	0.9	-935.897	-944.148	-936.290
0.5	1	-934.831	-938.919	-935.044
1	0.5	-942.159	-949.863	-947.533
1	0.75	-932.524	-939.676	-935.566
1	0.9	-934.214	-935.043	-936.026
1	1	-935.863	-946.046	-937.403
1.5	0.5	-935.541^{m}	-974.045	-936.795
1.5	0.75	-939.424	-974.058	-943.940
1.5	0.9	-938.154	-936.503	-927.700
1.5	1	-939.513	-949.609	-926.714
2	0.5	-935.507	-965.398	-938.099
2	0.75	-937.578	-972.612	-961.316
2	0.9	-937.221	-954.907	-928.153
2	1	-941.214	-964.301	-949.004

F, Cr, C and D indicate tuning parameter values leading to relevant starting points. The simulated log-likelihood at each solution is reported. Superscript m means the solution yields a higher $logL$ than the corresponding DE-assisted or PSO-assisted solution in Table A1. Stata classifies none of the solutions "converged", meaning that their Hessian (H) is not negative definite and/or their weighted gradient norm ($g'H^{-1}g$) exceeds -1E-5 in magnitude.

Table OA16. Pizza A: all DE+NM and PSO+NM solutions

A. DE+NM solutions

F	Cr	Start 1	Start 2	Start 3
0.2	0.2	-1436.89	-1441.79	-1411.76
0.2	0.4	-1403.72	-1439.35	-1398.61
0.2	0.6	-1428.65	-1409.77	-1405.73
0.2	0.8	-1397.52	-1413.04	-1404.69
0.4	0.2	-1370.37	-1370.68	-1370.54
0.4	0.4	-1368.72	-1380.03	-1373.67
0.4	0.6	-1367.75	-1366.81	-1368.16
0.4	0.8	-1367.22	-1365.86	-1366.39
0.6	0.2	-1370.59	-1364.64	-1367.64
0.6	0.4	-1366.74	-1365.98	-1369.10
0.6	0.6	-1366.31	-1364.07	-1367.71
0.6	0.8	-1362.46	-1363.00	-1369.52
0.8	0.2	-1370.07	-1367.66	-1367.96
0.8	0.4	-1368.77	-1366.68	-1372.47
0.8	0.6	-1375.61	-1375.71	-1368.60
0.8	0.8	-1372.64	-1373.97	-1367.68^{m}

B. PSO+NM solutions

C	D	Start 1	Start 2	Start 3
0.5	0.5	-1446.86	-1456.58	-1454.54
0.5	0.75	-1443.24	-1417.43	-1451.04
0.5	0.9	-1440.39	-1390.16	-1448.70
0.5	1	-1439.40	-1399.50	-1455.58
1	0.5	-1452.96	-1451.25	-1446.53
1	0.75	-1447.40	-1444.45	-1443.58
1	0.9	-1450.10	-1442.67	-1427.32
1	1	-1445.10	-1447.71	-1444.21
1.5	0.5	-1447.56	-1407.91	-1448.19
1.5	0.75	-1436.80	-1438.49	-1445.21
1.5	0.9	-1404.10	-1443.87	-1445.38
1.5	1	-1445.26	-1418.00	-1451.48
2	0.5	-1445.56	-1446.61	-1451.06
2	0.75	-1447.61	-1444.24	-1398.56
2	0.9	-1444.03	-1422.03	-1447.49
2	1	-1424.36	-1447.70	-1447.75

F, Cr, C and D indicate tuning parameter values leading to relevant starting points. The simulated log-likelihood at each solution is reported. Superscript m means the solution yields a higher $logL$ than the corresponding DE-assisted or PSO-assisted solution in Table A2. Stata classifies none of the solutions "converged", meaning that their Hessian (H) is not negative definite and/or their weighted gradient norm ($g'H^{-1}g$) exceeds -1E-5 in magnitude.

5.5 All data sets: mixed logit case studies

While our focus so far has been on GMNL, the presence of several local maxima is a feature of all random parameter logit (RPL) models. The DE- and PSO-assisted estimation strategies can be readily adapted to the estimation of other RPL models, and in this sub-section we explore whether our findings are generalizable to the standard mixed logit model (MIXL). For the four data sets in use, the preferred MIXL specification of Fiebig et al. (2010) constrains the off-diagonal elements of $\mathbf{\Sigma}$ to zero, like their preferred GMNL specification. We take their preferred MIXL specification as given and estimate the mean ($\boldsymbol{\beta}$) and standard deviations ($\boldsymbol{\sigma}$) of the normally distributed coefficients.

For each data set, several MIXL solutions have been obtained using the same tuning parameter values for the DE and PSO algorithms as in the previous sections. Only one conventional solution has been obtained in this case since using the MNL coefficients as starting values is likely to be the most common strategy for estimating MIXL. As far as we are aware, no previous study has made an explicit mention of starting values used in estimating the MIXL specification of interest here, presumably because the underlying optimization task may be perceived as numerically simple in that the postulated utility function is linear in parameters and convergence to local maxima can be achieved from a wide range of starting values.

Table OA17 reports the best ten (out of 48) DE-assisted solutions and best ten (out of 48) PSO-assisted solutions for Pap Smear, and Table OA18 reports the corresponding results for Pizza A. Table OA19 reports all DE- and PSO-assisted solutions for Holiday A, and Table OA20 report the corresponding results for Mobile Phone.

The results suggest that our earlier findings on the performance of the DE- and PSO-assisted strategies are not exclusively associated with GMNL. Despite the relative numerical simplicity of the MIXL optimization task, the DE- and PSO-assisted strategies perform better than the conventional strategy. The DE-assisted strategy still outperforms the PSO-assisted strategy in that the former locates solutions improving on the conventional solution with a greater frequency, and it also finds the best solution out of the ones we have obtained. Moreover, the DE-assisted results from the Pap Smear and Pizza A data sets show that configurations in $\texttt{F} = \{0.6, 0.8\} \times \texttt{Cr} = \{0.2, 0.4\}$ are well-suited to the MIXL specification too.

Tables OA21 to OA29 present the parameter estimates at the conventional and best DE-assisted solutions, and the statistics derived from those estimates. Tables OA21, OA22 and OA23 report the Pap Smear results. Tables OA24 and OA25 report the Pizza A results. Tables OA26 and OA27 report the Holiday A results. Finally, Tables OA28 and OA29 report the Mobile Phone results.

One notable difference when comparing the MIXL and GMNL results is that the MIXL solutions at various local maxima show much greater agreement in terms of policy-relevant statistics than the GMNL solutions do. Presumably because the random scale factor, which can influence all other parameters, is absent in MIXL, the coefficient estimates are very similar and produce almost the same percentiles of the WTP distributions.

Finally, we note that all conventional MIXL solutions display acceptable convergence diagnostics. In all data sets, including Pizza A, MIXL appears empirically identified as increasing the number of draws to 10,000 leads to similarly acceptable conventional solutions: the log-likelihood at the resulting conventional solution is -945.905 in Pap Smear, -1383.77 in Pizza A, -2545.08 in Holiday A, and -3971.01 in Mobile Phone.

Table OA17. Pap Smear: 10-best DE- and PSO-assisted mixed logit solutions

A. DE-assisted solutions

F	Cr	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.8	0.8	**-941.702**	1.02E-06	-1.23E-13	87.03318
0.8	0.6	**-941.702**	2.58E-05	-9.75E-12	87.03349
0.8	0.2	**-941.702**	4.06E-05	-3.96E-11	87.03226
0.8	0.4	**-942.609**	8.94E-08	-6.25E-16	60.53516
0.8	0.2	**-942.609**	6.63E-06	-3.07E-12	60.53556
0.8	0.6	**-942.609**	0.000392	-9.62E-09	60.53359
0.8	0.8	**-942.905**	2.97E-04	-8.87E-10	74.89967
0.8	0.4	**-943.098**	9.66E-05	-1.15E-09	88.7544
0.4	0.2	**-944.104**	1.10E-08	-6.39E-18	100.5098
0.4	0.2	**-944.104**	2.24E-08	-1.82E-17	100.5102

B. PSO-assisted solutions

C	D	$logL$	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.5	0.75	**-944.104**	3.60E-09	-7.33E-19	100.5097
1.5	0.75	**-944.104**	1.89E-08	-4.25E-18	100.51
1.5	1.00	**-944.104**	5.08E-08	-2.51E-17	100.5103
0.5	0.75	**-944.104**	5.66E-08	-2.17E-16	100.5106
0.5	1.00	**-944.104**	4.41E-07	-1.72E-15	100.5101
1.5	0.50	**-944.104**	5.82E-07	-2.75E-15	100.5094
1.5	0.75	**-944.104**	5.53E-07	-2.92E-15	100.5101
1	0.75	**-944.104**	1.11E-06	-1.04E-14	100.5105
0.5	0.50	**-944.104**	1.28E-06	-1.12E-13	100.5102
1	0.90	**-944.104**	2.00E-06	-1.20E-13	100.5095

The mixed logit model is the same as what has been specified to obtain the MIXL starting point. The conventional solution using the MNL starting point gives the simulated log-likelihood of -948.446. $logL$ is in boldface if it is greater than -948.446. The DE- and PSO-assisted strategies have been implemented with exactly the same computational settings as those used to estimate GMNL in Section 4 of the main text. See notes to Table OA4 for other information.

Table OA18. Pizza A: 10-best DE- and PSO-assisted mixed logit solutions

A. DE-assisted solutions

F	Cr	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.8	0.4	**-1380.49**	6.33E-09	-9.77E-19	15.06392
0.8	0.6	**-1380.49**	1.56E-08	-5.35E-18	15.06392
0.6	0.8	**-1380.49**	3.60E-08	-9.32E-18	15.06392
0.6	0.4	**-1380.49**	2.22E-05	-2.53E-12	15.06393
0.8	0.2	**-1380.49**	2.32E-05	-2.00E-11	15.06392
0.8	0.4	**-1381.21**	7.36E-10	-3.14E-20	15.02928
0.8	0.4	**-1381.21**	1.16E-08	-5.07E-18	15.02928
0.6	0.4	**-1381.21**	1.91E-05	-4.68E-12	15.0293
0.6	0.6	**-1381.21**	7.49E-05	-1.74E-10	15.02926
0.8	0.6	**-1381.48**	3.58E-09	-1.45E-19	14.77773

B. PSO-assisted solutions

C	D	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.5	0.5	**-1386.53**	4.58E-10	-4.31E-21	11.53264
1	0.75	**-1386.53**	3.62E-05	-1.69E-11	11.53262
1.5	0.5	**-1388.18**	1.08E-07	-3.01E-16	13.31187
0.5	0.9	**-1388.18**	0.000111	-1.30E-10	13.31188
2	0.75	**-1390.91**	6.32E-06	-8.16E-13	12.63712
2	0.5	**-1391.25**	2.27E-09	-1.07E-19	15.79458
0.5	0.5	**-1391.25**	2.38E-09	-1.38E-19	15.79457
0.5	0.5	**-1391.25**	7.17E-09	-1.86E-18	15.79457
1.5	0.9	**-1391.25**	3.01E-08	-6.82E-18	15.79455
1	1	**-1391.25**	4.89E-07	-6.05E-15	15.79458

The conventional solution using the MNL starting point gives the simulated log-likelihood of -1391.93. *logL* is in boldface if it is greater than -1391.93. See notes to Table OA17 for other information.

Table OA19. Holiday A: all 12 DE- and PSO-assisted mixed logit solutions

A. DE-assisted solutions

F	Cr	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.6	0.2	**-2545.35**	7.97E-07	-2.08E-15	9.503648
0.6	0.4	**-2545.35**	1.60E-06	-1.04E-14	9.503644
0.8	0.2	**-2545.35**	2.38E-06	-1.09E-13	9.503646
0.6	0.4	**-2545.35**	9.03E-06	-2.29E-13	9.503646
0.8	0.2	**-2545.35**	1.74E-05	-2.64E-12	9.503644
0.8	0.2	**-2545.35**	0.000106	-1.78E-10	9.503636
0.8	0.4	**-2545.35**	0.000374	-1.44E-09	9.503623
0.6	0.2	**-2551.18**	5.83E-08	-4.28E-17	9.530569
0.6	0.4	**-2551.18**	1.66E-07	-1.88E-16	9.530562
0.8	0.4	**-2552.16**	5.23E-08	-2.76E-17	9.847261
0.8	0.4	**-2553.92**	1.02E-09	-8.70E-21	8.942473
0.6	0.2	-2558.99	6.09E-07	-2.02E-15	9.997762

B. PSO-assisted solutions

C	D	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
1.5	0.75	**-2552.25**	2.00E-08	-4.04E-18	9.422702
1.5	0.75	**-2556.26**	4.87E-09	-1.17E-19	11.20861
2	0.75	**-2557.15**	1.36E-09	-3.32E-20	9.050988
2	0.9	**-2557.15**	3.34E-08	-3.96E-17	9.050987
1.5	0.75	**-2557.15**	8.92E-06	-3.93E-13	9.050994
2	0.75	**-2557.48**	1.19E-07	-6.65E-17	9.763996
1.5	0.9	**-2557.48**	1.15E-07	-2.29E-16	9.763993
2	0.9	**-2557.48**	2.15E-07	-3.98E-16	9.763992
2	0.9	-2559.46	1.73E-09	-3.69E-20	9.792831
2	0.75	-2560.24	0.000201	-5.62E-10	9.494974
1.5	0.9	-2560.61	1.06E-05	-1.18E-12	9.914803
1.5	0.9	-2569.68	3.76E-07	-5.88E-16	9.154277

The conventional solution using the MNL starting point gives the simulated log-likelihood of -2558.99. *logL* is in boldface if it is greater than -2558.99. See notes to Table OA17 for other information.

Table OA20. Mobile Phone: all 12 DE- and PSO-assisted mixed logit solutions

A. DE-assisted solutions

F	Cr	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
0.6	0.4	**-3968.55**	5.93E-06	-1.39E-12	93.71636
0.6	0.4	**-3968.55**	6.88E-05	-8.32E-11	93.69902
0.6	0.2	**-3968.55**	6.03E-05	-1.42E-10	93.73179
0.6	0.4	**-3968.55**	0.000264	-2.78E-09	93.71214
0.8	0.2	**-3968.55**	0.000548	-1.51E-08	93.72461
0.8	0.2	**-3968.55**	0.001697	-2.17E-07	93.62836
0.6	0.2	**-3968.55**	0.0028	-4.01E-07	93.48322
0.6	0.2	**-3968.55**	0.003644	-6.99E-07	93.42057
0.8	0.4	**-3968.6**	0.000034	-3.12E-10	390.5829
0.8	0.2	**-3968.6**	0.001789	-4.97E-08	389.8491
0.8	0.4	**-3968.6**	0.000674	-1.20E-07	388.0938
0.8	0.4	**-3968.6**	0.001047	-3.19E-07	387.1909

B. PSO-assisted solutions

C	D	*logL*	$\|g\|_\infty$	$g'H^{-1}g$	$\kappa(H)$
2	0.75	**-3968.55**	2.10E-06	-5.48E-14	93.73955
1.5	0.75	**-3968.55**	0.000522	-1.15E-08	93.68372
2	0.9	**-3968.55**	0.000781	-3.97E-08	93.66655
2	0.9	**-3968.6**	0.00117	-1.75E-08	390.1555
1.5	0.75	**-3968.6**	0.000603	-9.81E-08	388.6249
1.5	0.9	**-3968.6**	0.000627	-1.06E-07	388.5673
1.5	0.9	**-3969**	0.000852	-9.74E-09	83.73753
2	0.75	**-3969.3**	2.59E-06	-1.06E-13	70.62393
2	0.9	**-3969.3**	0.00014	-8.56E-10	70.62639
1.5	0.9	**-3969.3**	0.000596	-5.56E-09	70.62389
1.5	0.75	**-3969.3**	0.001086	-1.20E-08	70.62689
2	0.75	**-3969.35**	1.32E-05	-7.13E-12	203.0936

The conventional solution using the MNL starting point gives the simulated log-likelihood of -3973.9. *logL* is in boldface if it is greater than -3973.9. See notes to Table OA17 for other information.

Table OA21. Pap Smear: mixed logit parameter estimates

	A. Conventional		B. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.
If know doctor	0.868***	(0.226)	0.860***	(0.213)
	[1.692***]	(0.215)	[1.646***]	(0.207)
If doctor is male	-1.407***	(0.288)	-1.315***	(0.261)
	[2.265***]	(0.271)	[2.553***]	(0.298)
If test is due	3.126***	(0.312)	3.014***	(0.342)
	[2.351***]	(0.208)	[2.623***]	(0.248)
If doctor recommends	1.239***	(0.197)	1.272***	(0.226)
	[0.887***]	(0.182)	[1.078***]	(0.144)
Test cost	-0.212***	(0.070)	-0.225***	(0.069)
	[0.133*]	(0.070)	[0.077]	(0.081)
ASC for test	-0.988***	(0.317)	-1.163***	(0.311)
	[2.927***]	(0.316)	[2.665***]	(0.284)
logL	-948.446		-941.702	

For each named attribute, the corresponding elements of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ (in [.]) are reported. *, **, *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Table OA22. Pap Smear: simulated WTP distributions (from mixed logit)

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
If know doctor:					
Conventional	-89	-18	35	101	206
Best DE-assisted	-60	-10	40	92	156
If doctor is male:					
Conventional	-306	-146	-58	14	111
Best DE-assisted	-237	-139	-57	19	92
If test is due:					
Conventional	-33	49	129	244	471
Best DE-assisted	-13	56	135	231	348
If doctor recommends:					
Conventional	-10	21	51	98	192
Best DE-assisted	-6	22	56	96	145

Figures are in $s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in Table OA21. p(Q) denotes the Q^{th} percentile of the simulated distribution.

Table OA23. Pap Smear: predicted choice probabilities (from mixed logit)

	A. Conventional	B. Best DE-assisted
Base choice probability	0.57	0.56
Change when test is not due	-0.26	-0.26
Change when don't know doctor	-0.07	-0.07
Change when doctor is female	+0.13	+0.12
Change when doctor recommends	+0.09	+0.09
Change when test cost-free	+0.05	+0.05

The probabilities have been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in Table OA21. The base choice probability is the probability of choosing a test (over no test) when the test is due, the patient knows the doctor, the doctor is male, the doctor makes no recommendation, and the cost is $30. Each row reports how this probability changes when each attribute changes from its base level.

Table OA24. Pizza A: mixed logit parameter estimates

	A. Conventional		B. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.
Gourmet	0.050	(0.054)	-0.016	(0.057)
	[0.533***]	(0.063)	[0.602***]	(0.065)
Price	-0.374***	(0.059)	-0.348***	(0.054)
	[0.560***]	(0.069)	[0.514***]	(0.057)
Ingredient freshness	1.076***	(0.106)	1.100***	(0.104)
	[1.077***]	(0.101)	[1.228***]	(0.115)
Delivery time	0.212***	(0.050)	0.233***	(0.054)
	[0.380***]	(0.067)	[0.455***]	(0.069)
Crust	0.111	(0.068)	0.103	(0.069)
	[0.812***]	(0.082)	[0.879***]	(0.089)
Sizes	0.191***	(0.054)	0.169***	(0.053)
	[0.481***]	(0.066)	[0.452***]	(0.061)
Steaming hot	0.772***	(0.084)	0.891***	(0.094)
	[1.036***]	(0.103)	[0.902***]	(0.083)
Late open hours	0.063	(0.042)	0.090**	(0.044)
	[0.295***]	(0.053)	[0.348***]	(0.061)
logL	-1391.929		-1380.492	

See notes to Table OA21.

Table OA25. Pizza A: simulated WTP distributions (from mixed logit)

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
Gourmet:					
Conventional	-9	-3	0	3	9
Best DE-assisted	-12	-4	0	4	11
Ingredient freshness:					
Conventional	-22	-3	5	12	30
Best DE-assisted	-25	-4	5	14	36
Delivery time:					
Conventional	-7	-2	1	3	9
Best DE-assisted	-9	-2	1	4	11
Crust:					
Conventional	-14	-4	0	5	15
Best DE-assisted	-16	-5	0	6	18
Sizes:					
Conventional	-8	-2	1	4	10
Best DE-assisted	-8	-2	1	4	10
Steaming hot:					
Conventional	-20	-4	3	11	27
Best DE-assisted	-20	-3	4	12	30
Late open hours:					
Conventional	-5	-1	0	2	6
Best DE-assisted	-6	-2	0	3	8

Figures are in $s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in Table OA24. p(Q) denotes the Q^{th} percentile of the simulated distribution.

Table OA26. Holiday A: mixed logit parameter estimates

	A. Conventional		B. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.
Price	-0.332***	(0.035)	-0.321***	(0.036)
	[0.334***]	(0.043)	[0.373***]	(0.045)
Overseas destination	0.196***	(0.067)	0.227***	(0.067)
	[1.127***]	(0.077)	[1.133***]	(0.073)
Airline	-0.029	(0.032)	-0.025	(0.033)
	[0.293***]	(0.045)	[0.306***]	(0.045)
Length of stay	0.521***	(0.045)	0.538***	(0.045)
	[0.589***]	(0.053)	[0.616***]	(0.053)
Meal inclusion	0.530***	(0.042)	0.555***	(0.044)
	[0.453***]	(0.048)	[0.495***]	(0.048)
Local tours availability	0.169***	(0.035)	0.188***	(0.035)
	[0.319***]	(0.051)	[0.308***]	(0.051)
Peak season	0.061*	(0.033)	0.061*	(0.035)
	[0.275***]	(0.052)	[0.314***]	(0.048)
4-star accommodation	0.840***	(0.055)	0.859***	(0.055)
	[0.793***]	(0.057)	[0.794***]	(0.057)
logL	-2558.989		-2545.354	

See notes to Table OA21.

Table OA27. Holiday A: simulated WTP distributions (from mixed logit)

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
Overseas destination:					
Conventional	-1249	-371	67	541	1484
Best DE-assisted	-1214	-363	57	512	1394
Airline:					
Conventional	-374	-133	-13	103	320
Best DE-assisted	-364	-127	-10	105	339
Length of stay:					
Conventional	-325	-63	60	209	527
Best DE-assisted	-319	-64	60	195	472
Meal inclusion:					
Conventional	-575	9	204	480	1081
Best DE-assisted	-609	-35	186	457	1069
Local tours availability:					
Conventional	-325	-63	60	209	527
Best DE-assisted	-319	-64	60	195	472
Peak season:					
Conventional	-312	-87	22	140	364
Best DE-assisted	-333	-100	18	144	396
4-star accommodation:					
Conventional	-917	-18	319	763	1767
Best DE-assisted	-1009	-61	286	696	1587

Figures are in \$s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in table OA26. p(Q) denotes the Q^{th} percentile of the simulated distribution.

Table OA28. Mobile Phone: mixed logit parameter estimates

	A. Conventional		B. Best DE-assisted	
	Estimate	Std. Err.	Estimate	Std. Err.
No voice comm.	0.042	(0.049)	0.037	(0.049)
	[0.195]	(0.126)	[0.219**]	(0.110)
Voice dialing	0.096**	(0.048)	0.096**	(0.048)
	[0.154]	(0.145)	[0.158]	(0.144)
Voice operation	-0.127**	(0.050)	-0.122**	(0.050)
	[0.127]	(0.150)	[0.081]	(0.211)
No push to com.	0.053	(0.047)	0.052	(0.047)
	[0.008]	(0.135)	[0.049]	(0.123)
Push to talk	0.047	(0.049)	0.054	(0.048)
	[0.231**]	(0.108)	[0.220**]	(0.098)
Push to share pics/video	-0.023	(0.048)	-0.026	(0.048)
	[0.080]	(0.157)	[0.005]	(0.196)
Personal e-mail	-0.074	(0.049)	-0.076	(0.049)
	[0.009]	(0.088)	[0.016]	(0.097)
Corporate e-mail	0.076	(0.047)	0.078*	(0.047)
	[0.110]	(0.141)	[0.066]	(0.230)
Both e-mails	-0.029	(0.048)	-0.034	(0.049)
	[0.114]	(0.156)	[0.165]	(0.104)
WiFi	-0.002	(0.027)	-0.002	(0.027)
	[0.018]	(0.062)	[0.024]	(0.068)
USB calbe/cradle	0.071**	(0.027)	0.070**	(0.028)
	[0.009]	(0.093)	[0.050]	(0.083)
Thermometer	0.069**	(0.028)	0.067**	(0.029)
	[0.164**]	(0.070)	[0.180***]	(0.068)
Flashlight	0.052*	(0.027)	0.049*	(0.028)
	[0.033]	(0.076)	[0.058]	(0.081)
Price/100	-0.760***	(0.062)	-0.778***	(0.062)
	[0.771***]	(0.066)	[0.764***]	(0.060)
ASC for purchase	-0.470***	(0.118)	-0.472***	(0.116)
	[1.932***]	(0.121)	[1.945***]	(0.119)
logL	-3973.897		-3968.545	

See notes to Table OA21.

Table OA29. Mobile Phone: simulated WTP distributions (from mixed logit)

Willingness-to-pay for	p(10)	p(25)	p(50)	p(75)	p(90)
No voice comm.:					
Conventional	-45	-14	4	21	56
Best DE-assisted	-47	-15	3	22	58
Voice dialing:					
Conventional	-33	-6	7	23	55
Best DE-assisted	-31	-6	7	23	54
Voice operation:					
Conventional	-59	-26	-11	0	29
Best DE-assisted	-46	-21	-10	-3	22
No push to com.:					
Conventional	-12	3	5	9	21
Best DE-assisted	-10	0	4	10	22
Push to talk:					
Conventional	-54	-17	4	26	66
Best DE-assisted	-48	-14	4	24	61
Push to share pics/video:					
Conventional	-25	-9	-2	5	18
Best DE-assisted	-9	-4	-2	-1	5
Personal e-mail:					
Conventional	-29	-13	-7	-4	16
Best DE-assisted	-27	-13	-7	-4	15
Corporate e-mail:					
Conventional	-23	-4	6	18	42
Best DE-assisted	-15	1	6	14	33
Both e-mails:					
Conventional	-34	-13	-2	7	27
Best DE-assisted	-43	-17	-3	11	37
WiFi:					
Conventional	-5	-2	-0	1	4
Best DE-assisted	-6	-2	-0	2	6
USB cable/cradle:					
Conventional	-15	4	7	12	28
Best DE-assisted	-13	1	6	13	28
Thermometer:					
Conventional	-35	-8	6	22	52
Best DE-assisted	-36	-9	5	22	53
Flashlight:					
Conventional	-10	1	4	9	21
Best DE-assisted	-11	-1	4	10	23

Figures are in $s. Each willingness-to-pay (WTP) distribution has been simulated by making 100,000 draws from the joint density of utility coefficients according to the solutions in OA28. p(Q) denotes the Q^{th} percentile of the simulated distribution.

Additional references

[1] Baudin M. 2-1-. Nelder-Mead User's Manual. `http://forge.scilab.org/index.php/p/docneldermead/downloads/98/`. *Last accessed at:* 14/04/2016.

[2] Bhat, C. 1997. An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science* **31**: 34-48.

[3] Hole, A. 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* **7**: 388-401.

[4] Liu, A., Yang, M-T. 2012. A New Hybrid Nelder-Mead Particle Swarm Optimization for Coordination Optimization of Directional Overcurrent Relays. *Mathematical Problems in Engineering*. doi:10.1155/2012/456047.

[5] Luchi, F., Krohling, RA. 2015. Differential Evolution and Nelder-Mead for constrained non-linear integer optimization problems. *Procedia Computer Science* **55**: 668-677.